



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

PROGRAMA DE DOCTORADO EN
TECNOLOGÍAS PARA LA SALUD Y EL BIENESTAR

Archetype development and governance methodologies for the electronic health record

DOCTORAL THESIS

Author:

David Moner Cano

Supervisors:

Dra. Montserrat Robles Viejo

Dr. José Alberto Maldonado Segura

February 2021

Acknowledgements

I would like to thank the following people for helping, not only in the development of this thesis, but in my professional career. Please accept my gratitude, irrespective of your name appearing in the acknowledgment.

I would like to express my sincere gratitude to Montserrat Robles and to José Alberto Maldonado, who gave me the opportunity to enter in the exciting world of medical informatics 15 years ago. I started with them, and they have always been by my side until now, as supervisors of this thesis.

To Diego Boscá, the person who has made possible that many dreams became a reality. He is possibly the person with whom I have spent more hours talking and sharing ideas, and I am sure it will continue being like this for many years.

To all my current and former colleagues at the Universitat Politècnica de Valencia and in VeraTech for Health. This would be a very long list of names of people with whom, at some point, I shared work, worries and celebrations.

To all the excellent professionals from all over the world who I had the opportunity to meet, discuss, learn, and many times share a pint of beer.

And last, but not least, to my family, who have always supported and trusted in me, even if they did not understand what my work was exactly about.

This thesis was partially funded by the *Ministerio de Economía y Competitividad, ayudas para contratos para la formación de doctores en empresas "Doctorados Industriales"*, grant DI-14-06564.

This thesis was partially funded by the *Agencia Valenciana de la Innovación, ayudas del Programa de Promoción del Talento – Doctorados empresariales (INNO-DOCTO)*, grant INNTA3/2020/12.

Contents

Acknowledgements	3
Contents	5
List of figures.....	9
List of tables.....	11
Abstract	13
Resumen.....	15
Resum.....	17
Abbreviations and acronyms.....	19
Chapter 1. Introduction.....	21
1.1. Motivation	21
1.2. Hypothesis, research questions and objectives.....	23
1.2.1. Hypothesis	23
1.2.2. Specific research questions and objectives	23
1.3. Contributions	24
1.3.1. Main contributions	24
1.3.2. Journal publications and conference contributions	26
1.3.3. Research visits.....	29
1.3.4. Grants.....	30
1.3.5. Projects	31
1.3.6. Software.....	32
1.3.7. Other contributions.....	32
1.4. Structure of the thesis.....	33

Chapter 2. Electronic health record formalization and semantic interoperability	35
2.1. Health information systems and electronic health records.....	35
2.2. Semantic interoperability of health data.....	39
2.3. Electronic health record standards	41
2.4. Clinical information models	45
2.5. Clinical terminologies	48
Chapter 3. Study of archetype modeling methodologies and governance initiatives	53
3.1. Modeling of archetypes	53
3.1.1. Software development methodologies.....	56
3.1.2. Clinical information modeling methodologies	56
3.1.3. Ontological modeling methodologies.....	57
3.2. Governance of archetypes.....	58
3.2.1. Governance initiatives and systems	61
3.2.2. Comparative analysis.....	69
Chapter 4. Archetype modeling methodology	77
4.1. Requirements.....	77
4.2. Work group.....	79
4.3. Selection of the Reference Model	80
4.4. Design guidelines.....	81
4.5. Methodology phases.....	81
4.5.1. Phase 1. Analysis	83
4.5.2. Phase 2. Design.....	85
4.5.3. Phase 3. Development.....	87

4.5.4. Phase 4. Validation	90
4.5.5. Phase 5. Publication	92
4.6. Discussion.....	93
Chapter 5. Archetype governance.....	99
5.1. Definition and scope.....	99
5.2. Management of participants and roles	102
5.3. Governance style.....	104
5.4. Archetype development strategy	107
5.5. Quality assurance	109
5.6. Publication and visibility	111
5.7. Alignment of archetype governance activities to PRINCE2 project management processes.....	112
Chapter 6. Software development and use cases.....	119
6.1. Software developments	119
6.1.1. LinkEHR Studio.....	119
6.1.2. LinkEHR Model Manager	140
6.2. Use cases	151
Chapter 7. Conclusions and future work.....	157
Bibliography.....	163

List of figures

Figure 1. Specializations of information systems in the health domain	36
Figure 2. Example of a standard reference model: main classes of ISO 13606.....	44
Figure 3. Summary of Reference Models and their Clinical Information Model definition artifacts, extracted from [11]	46
Figure 4. Example of a patient summary archetype	47
Figure 5. Areas of knowledge governance.....	59
Figure 6. openEHR international CKM	62
Figure 7. Number of archetypes in the openEHR CKM as of June 2020.....	63
Figure 8. Number of templates in the openEHR CKM as of June 2020.....	63
Figure 9. Healthcare Modeling Collaboration archetype manager	64
Figure 10. ArchMS system	65
Figure 11. ART-DECOR management system.....	66
Figure 12. Example of a SIMPLIFIER.NET project	67
Figure 13. SOM-SNS governance platform	68
Figure 14. Graphic summary of the Archetype Modeling Methodology workflow	82
Figure 15. A mind map can easily represent the set of concepts and information elements discovered during the analysis phase	83
Figure 16. Example of a design table created during Phase 2	86
Figure 17. Information structure of a medication archetype implemented in ISO 13606.....	87
Figure 18. Workflow of PRINCE2 processes.....	115
Figure 19. Example of an archetype expressed in ADL	120
Figure 20. OpenEHR Archetype Editor	122
Figure 21. OpenEHR Template Designer.....	123
Figure 22. Archetype Designer.....	124
Figure 23. LinkEHR Studio.....	126
Figure 24. Import of a new RM in LinkEHR Studio	127

Figure 25. Generic archetype tree representing an Allergy archetype in HL7 CDA	128
Figure 26. A specific HL7 CDA archetype tree representing an Allergy archetype	129
Figure 27. Edition of the documentation of a reference model.....	130
Figure 28. openEHR medication use archetype	132
Figure 29. Archetype-based formulary.....	132
Figure 30. Archetype-based mind map.....	133
Figure 31. Archetype-based table summary.....	134
Figure 32. Archetype-based implementation guide	135
Figure 33. Integration of LinkEHR Studio with archetype repositories	136
Figure 34. Example of HL7 FHIR ValueSets.....	137
Figure 35. Access to the semantic binding of archetypes.....	138
Figure 36. Search windows of terminology concepts and ValueSets.....	139
Figure 37. Interface of LinkEHR Model Manager.....	142
Figure 38. LinkEHR Model Manager modules.....	143
Figure 39. Example relationships between clinical models	146
Figure 40. Visualization of model relationships in a mind map.....	147
Figure 41. Reference set definition in LinkEHR Model Manager.....	149
Figure 42. Edition of role permissions.....	150

List of tables

Table 1. Structure of the thesis and relationship with the research questions, objectives and contributions.....	34
Table 2. Summary of modeling methodologies and their domain of use.....	55
Table 3. Comparative analysis of CIM governance systems	72
Table 4. Comparison of AMM, CIM, software and ontology development methodologies.....	95
Table 5. Types of governance styles and applicable characteristics	104
Table 6. Correspondence between PRINCE2 processes and governance activities	116

Abstract

Semantic interoperability of health information is an essential requirement for the sustainability of healthcare, and it is essential to face the new health challenges of a globalized world. This thesis provides new methodologies to tackle some of the fundamental aspects of semantic interoperability, specifically those aspects related to the definition and governance of clinical information models expressed in the form of archetypes.

The contributions of the thesis are:

- Study of existing modeling methodologies of semantic interoperability components that will influence in the definition of an archetype modeling methodology.
- Comparative analysis of existing clinical information model governance systems and initiatives.
- A proposal of a unified Archetype Modeling Methodology that formalizes the phases of archetype development, the required participants, and the good practices to be followed.
- Identification and definition of archetype governance principles and characteristics.
- Design and development of tools that provide support to archetype modeling and governance.

The contributions of this thesis have been put into practice in multiple projects and development experiences. These experiences vary from a local project inside a single organization that required a reuse on clinical data based on semantic interoperability principles, to the development of national electronic health record projects.

Resumen

La interoperabilidad semántica de la información sanitaria es un requisito imprescindible para la sostenibilidad de la atención sanitaria, y es fundamental para afrontar los nuevos retos sanitarios de un mundo globalizado. Esta tesis aporta nuevas metodologías para abordar algunos de los aspectos fundamentales de la interoperabilidad semántica, específicamente aquellos relacionados con la definición y gobernanza de modelos de información clínica expresados en forma de arquetipo.

Las aportaciones de la tesis son:

- Estudio de las metodologías de modelado existentes de componentes de interoperabilidad semántica que influirán en la definición de una metodología de modelado de arquetipos.
- Análisis comparativo de los sistemas e iniciativas existentes para la gobernanza de modelos de información clínica.
- Una propuesta de Metodología de Modelado de Arquetipos unificada que formalice las fases de desarrollo del arquetipo, los participantes requeridos y las buenas prácticas a seguir.
- Identificación y definición de principios y características de gobernanza de arquetipos.
- Diseño y desarrollo de herramientas que brinden soporte al modelado y la gobernanza de arquetipos.

Las aportaciones de esta tesis se han puesto en práctica en múltiples proyectos y experiencias de desarrollo. Estas experiencias varían desde un proyecto local dentro de una sola organización que requirió la reutilización de datos clínicos basados en principios de interoperabilidad semántica, hasta el desarrollo de proyectos de historia clínica electrónica de alcance nacional.

Resum

La interoperabilitat semàntica de la informació sanitària és un requisit imprescindible per a la sostenibilitat de l'atenció sanitària, i és fonamental per a afrontar els nous reptes sanitaris d'un món globalitzat. Aquesta tesi aporta noves metodologies per a abordar alguns dels aspectes fonamentals de la interoperabilitat semàntica, específicament aquells relacionats amb la definició i govern de models d'informació clínica expressats en forma d'arquetip.

Les aportacions de la tesi són:

- Estudi de les metodologies de modelatge existents de components d'interoperabilitat semàntica que influiran en la definició d'una metodologia de modelatge d'arquetips.
- Anàlisi comparativa dels sistemes i iniciatives existents per al govern de models d'informació clínica.
- Una proposta de Metodologia de Modelatge d'Arquetips unificada que formalitza les fases de desenvolupament de l'arquetip, els participants requerits i les bones pràctiques a seguir.
- Identificació i definició de principis i característiques de govern d'arquetips.
- Disseny i desenvolupament d'eines que brinden suport al modelatge i al govern d'arquetips.

Les aportacions d'aquesta tesi s'han posat en pràctica en múltiples projectes i experiències de desenvolupament. Aquestes experiències varien des d'un projecte local dins d'una sola organització que va requerir la reutilització de dades clíniques basades en principis d'interoperabilitat semàntica, fins al desenvolupament de projectes d'història clínica electrònica d'abast nacional.

Abbreviations and acronyms

ADL	Archetype Definition Language
AMM	Archetype Modeling Methodology
API	Application Programming Interface
ASTM	American Society for Testing and Materials
CCD	Continuity of Care Document
CCR	Continuity of Care Record
CDA	Clinical Document Architecture
CDISC	Clinical Data Interchange Standards Consortium
CEM	Clinical Element Model
CIM	Clinical Information Model
CIMI	Clinical Information Modeling Initiative
CKG	Clinical Knowledge Governance
CKM	Clinical Knowledge Manager
DCM	Detailed Clinical Model
EHR	Electronic Health Record
FHIR	Fast Healthcare Interoperability Resources
HL7	Health Level Seven
HMC	Healthcare Modeling Collaboration
IT	Information Technologies
MML	Medical Markup Language
ODM	Operational Data Model

OWL	Web Ontology Language
RIM	Reference Information Model
RM	Reference Model
UN	United Nations
WHO	World Health Organization

Chapter 1

Introduction

1.1. Motivation

The formalization and standardization of medical data contained in the electronic health record (EHR) has been a major objective of medical informatics during the last decades. The aim of these efforts has been to improve the availability and quality of health data registered in EHR systems. This has a direct impact on healthcare provided to the patient, but also on the possibility of safely reusing EHRs as a source of information for clinical research, population health, or healthcare management.

The ability of reusing EHR data is part of the broader problem of semantic interoperability. Semantic interoperability is the ability of information systems to seamlessly share, analyze, aggregate and understand information and data that has been generated by other systems [1]. This is especially important in the case of EHRs, where clinical data is usually distributed among different information systems or databases. To build semantically interoperable health information systems, it is essential to describe formally the data contained in the EHR, so that it becomes meaningful information that can be processed and exchanged.

Semantic interoperability relies upon three layers of artefacts to represent meaning [1]: standard reference models, standard terminologies and standard clinical information models.

A Reference Model (RM) defines the generic classes and attributes that can represent any clinical data of the EHR together with its contextual information, needed to ensure the correct future interpretation of data.

Medical terminologies provide the specific and univocal meaning of concepts and terms used in healthcare and inside EHRs, to ensure the correct interpretation of data.

Finally, clinical information models define specific structures of data by constraining the generic properties of a RM, and, at the same time, they incorporate terminology mappings to provide a semantic definition of those structures. In this thesis, archetypes will be the means of representation of clinical information models.

Defining archetypes for local use in a controlled scenario is relatively easy. However, defining archetypes by agreement of an international community that cover different use cases and needs is a complex task. We face centuries of tradition and practice in medical documentation using paper and free text, where the primary recipients of medical information were clinicians or the patients themselves. Nowadays the recipients are mainly EHR systems that automatically process the clinical information and can act or take decisions over it. It is necessary to create information models based on an interdisciplinary work and collaboration between health professionals and information technology professionals.

The author of this thesis has faced these difficulties for over 15 years. First, in the Biomedical Informatics Group (IBIME) of the Institute of Information and Communication Technologies (ITACA) of the Universitat Politècnica de València (UPV). Afterwards, in VeraTech for Health, a spin-off company of UPV. In both organizations, the objectives have remained the same: to develop solutions that facilitate the standardization and semantic interoperability of EHRs. This thesis arises from this experience and pursues the following research questions and objectives.

1.2. Hypothesis, research questions and objectives

1.2.1. Hypothesis

The complete standardization and semantic interoperability of medical information lacks a detailed methodology that guides the development and governance of archetypes. It is possible to formalize an archetype modeling methodology and to specify the principles of archetype governance.

1.2.2. Specific research questions and objectives

In order to confirm the hypothesis, the following research questions are identified:

- **R1.** Can the process of definition and implementation of archetypes be formalized?
- **R2.** Can guidelines and best practices for building archetypes be provided?
- **R3.** Can governance rules be defined to simplify the development and deployment of archetypes?
- **R4.** Can tools that support archetype modeling and governance incorporate guidance to facilitate those activities?

The following research objectives will answer satisfactorily the previous research questions:

- **O1.** To analyze current archetype modeling and governance methodologies.
- **O2.** To formalize and improve archetype modeling processes.
- **O3.** To formalize and improve archetype governance processes.
- **O4.** To provide tools that support archetype modeling and governance.

1.3. Contributions

1.3.1. Main contributions

- **C1. Study of clinical information model development methodologies.**

Contributes to research objective O1.

A systematic review of development methodologies of clinical information models was performed. The systematic review aimed to identify and compare the existing processes and methodologies that have been published in the literature for defining clinical information models that support the semantic interoperability of Electronic Health Record systems. 378 articles were screened and 36 papers were selected for full review. The systematic review confirmed that most of the clinical information modeling activities follow very similar steps for the definition of clinical information models. However, they lack a formal and homogeneous definition of the development methodologies followed. This study was published in A. Moreno-Conde, D. Moner, W.D. da Cruz, M.R. Santos, J.A. Maldonado, M. Robles, D. Kalra, *Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis*, Journal of the American Medical Informatics Association. (2015). doi:10.1093/jamia/ocv008.

- **C2. Formalization of an archetype modeling methodology.**

Contributes to research objective O2.

An archetype modeling methodology has been developed. This methodology is inspired by existing best practices of clinical information models, software, and ontology development. The methodology is composed by five phases, namely analysis, design, development, validation, and publication. It helps to the definition and adoption of interoperable archetypes, improves their quality, and facilitates their reuse among different information systems and EHR projects. Moreover, the proposed methodology can be also a reference for the development of any other clinical information model formalism other than archetypes. The methodology has been applied and evaluated in regional and national EHR

projects. The complete methodology was published in D. Moner, J.A. Maldonado, M. Robles, *Archetype modeling methodology*, Journal of Biomedical Informatics. 79 (2018) 71–81. doi:10.1016/j.jbi.2018.02.003.

- **C3. Analysis of archetype governance experiences and elaboration of governance requirements.**

Contributes to research objective O1 and O3.

Archetype governance includes the set of policies, actions, and tools to ensure that archetypes can be identified, stored, searched, and that they can evolve to meet new requirements. An analysis of existing experiences of archetype governance has been carried out, including the governance systems of relevant international EHR projects based on archetypes. This study served to identify best practices in archetype governance and to obtain the requirements of tools that support that governance process. The results of this analysis can be found in Chapter 5.

- **C4. Development of tools to facilitate archetype modeling and governance.**

Contributes to research objective O4.

Several software solutions for archetype development have been developed. These solutions support different phases of the archetype modeling methodology defined in C2. Design phase is supported by the generation of visual representations of the contents of the archetype, such as mind maps and summary tables. They are useful for health domain experts to discuss the contents and structure of the developed archetypes. For the development phase, a reference-model independent archetype editor has been developed. This tool supports all the requirements to develop archetypes. Finally, the validation phase is supported by including automatic validation mechanisms of the developed archetypes. Also, by generating additional resources such as implementation guides and sample data instances conformant to archetypes. They help technical implementers in the development and testing of the generated archetypes.

These solutions have been incorporated as part of LinkEHR Studio, a multi-model archetype editor described in Section 6.1.1. They have been also published in multiple of the contributions listed in next section, including the publications P2, P5, P6, P7, P9, P11 and P12.

- **C5. Design and development of a clinical information model management system.**

Contributes to research objective O4.

A clinical information model management system aligned to best practices identified in C3 has been developed. This system allows the management of different types of interoperability resources including archetypes, implementation guides, templates and any other documentation. A characteristic of this system is that it manages the concept of clinical information model as the center of all resources. A person in charge of the management can define semantic relationships between clinical information models and attach any type of interoperability resource to them. The management systems, called LinkEHR Model Manager, is described in Section 6.1.2. It has been also published in P4 and P10.

1.3.2. Journal publications and conference contributions

The work presented in this thesis spreads through over a decade of research and development. The following journal publications and conference contributions describe the results achieved, and reflect the work done in the context of specific research projects that required the modelling and/or governance of archetypes or their use in practical implementations and projects.

Journal publications

- **P1. D. Moner, J.A. Maldonado, M. Robles, Archetype modeling methodology, Journal of Biomedical Informatics. 79 (2018) 71–81. doi:10.1016/j.jbi.2018.02.003.** (Q1, 2018) This publication directly contributes to the results of this thesis. It describes the archetype modeling methodology included in Chapter 4 of the thesis.

- **P2.** L. Marco-Ruiz, **D. Moner**, J.A. Maldonado, N. Kolstrup, J.G. Bellika, *Archetype-based data warehouse environment to enable the reuse of electronic health record data*, *International Journal of Medical Informatics*. 84 (2015) 702–714. doi:10.1016/j.ijmedinf.2015.05.016. (Q1, 2015).
- **P3.** A. Moreno-Conde, **D. Moner**, W.D. da Cruz, M.R. Santos, J.A. Maldonado, M. Robles, D. Kalra, *Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis*, *Journal of the American Medical Informatics Association*. (2015). doi:10.1093/jamia/ocv008. (Q1, 2015).

Conference contributions

- **P4.** D. Bosca, L. Marco, **D. Moner**, J.A. Maldonado, L. Insa, M. Robles, *Detailed Clinical Models Governance System in a Regional EHR Project*, in: L.M. Roa Romero (Ed.), *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, Springer International Publishing, Cham, 2014: pp. 1266–1269. http://link.springer.com/10.1007/978-3-319-00846-2_313.

This publication directly contributes to the results of this thesis. It describes the first studies and developments about archetype governance included in Chapter 5.

- **P5.** C. Sáez, **D. Moner**, R. García-De-León-Chocano, V. Muñoz-Soler, R. García-De-León-González, J.A. Maldonado, D. Boscá, S. Tortajada, M. Robles, J.M. García-Gómez, M. Alcaraz, P. Serrano, J.L. Bernal, J. Rodríguez, G. Bustos, M. Esparza. *A Standardized and Data Quality Assessed Maternal-Child Care Integrated Data Repository for Research and Monitoring of Best Practices: A Pilot Project in Spain*, *Informatics for Health Congress 2017*, in: *Studies in Health Technology and Informatics*. 539–543. doi:10.3233/978-1-61499-753-5-539.
- **P6.** **D. Moner**, M. Terrón, C. Angulo, L. Lechuga, P. Serrano-Balazote, J.A. Maldonado, F.J. Farfán, M. Robles, *Implementation of a CEN/ISO 13606 Platform for Medicines Reconciliation*, in: *Proceedings MIE 2011*, Oslo, 2011. doi:10.13140/2.1.1894.7846.

- **P7. D. Moner**, J.A. Maldonado, C. Angulo, D. Bosca, D. Perez, I. Abad, E. Reig, M. Robles, *Standardization of discharge reports with the ISO 13606 norm*, in: *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, IEEE, Vancouver, 2008*: pp. 1470–1473. doi:10.1109/IEMBS.2008.4649445.
- **P8. D. Moner**, J.A. Maldonado, D. Boscá, C. Angulo, M. Robles, D. Pérez, P. Serrano, *CEN EN13606 normalisation framework implementation experiences*, *EFMI Special Topic Conference 2010*, in: *Stud Health Technol Inform.* 155 (2010) 136–142.
- **P9. D. Moner**, A. Moreno, J.A. Maldonado, M. Robles, C. Parra. *Using archetypes for defining CDA templates*, *XXIVth International Congress of the European Federation for Medical Informatics (MIE 2012)*, in: *Stud Health Technol Inform.* 180 (2012) 53–57.
- **P10. J.T. Fernandez-Breis**, M. Menarguez-Tortosa, C. Martinez-Costa, E. Fernandez-Breis, J. Herrero-Sempere, **D. Moner**, J. Sanchez, R. Valencia-Garcia, M. Robles, *A Semantic Web-based System for Managing Clinical Archetypes*, *Conf Proc IEEE Eng Med Biol Soc. 2008 (2008)* 1482–1485. doi:10.1109/IEMBS.2008.4649448.
- **P11. D. Moner**, J.A. Maldonado, D. Bosca, J.T. Fernandez, C. Angulo, P. Crespo, P.J. Vivancos, M. Robles, *Archetype-Based Semantic Integration and Standardization of Clinical Data*, in: *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, IEEE, New York, 2006*: pp. 5141–5144. doi:10.1109/IEMBS.2006.260124.
- **P12. D. Moner**, J.A. Maldonado, D. Boscá, A. Mañas, M. Robles, *Development of a Visual Editor for the Definition of HL7 CDA Archetypes*, in: L.M. Roa Romero (Ed.), *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, Springer International Publishing, Cham, 2014: pp. 1258–1261. http://link.springer.com/10.1007/978-3-319-00846-2_311.

P2, P3, P5, P6, P7 and P8 describe the development of archetypes in different domains. All these developments helped in testing and improving the archetype modeling methodology, published in P1.

P4, P8 and P10 describe experiences where a management of archetypes took place.

P4, P2, P5, P6, P7, P9, P11 and P12 describe tools and implementations that facilitate the archetype definition and management processes.

1.3.3. Research visits

The author of the thesis did a research visit to Norway in 2014. He applied some of the initial results of this thesis during the visit. In particular, an initial version of the archetype modeling methodology was used to develop archetypes for clinical research.

- **Institution:** Norwegian Centre for Integrated Care and Telemedicine, Tromsø, Norway.
- **Date:** From April 14th 2014 to June 16th 2014.
- **Objective:** The main objective of the visit was to develop a solution that allows clinical users to define standard archetype-based queries to retrieve clinical data from existing electronic health record systems. Data retrieved can be exchanged among health information systems or be reused for clinical research or public health studies. This reuse of data improves the sustainability of health services by controlling costs by avoiding the repetition of clinical tests and improves the quality of healthcare provided to patients.
- **Results:** A software platform for the reuse of clinical data based on archetypes was designed and implemented. This platform covers all the needed steps in a clinical research project from the IT perspective: data acquisition, data normalization, data storage and data query. One of the main results of this work was the proposal of the METL methodology for data integration and normalization. METL stands for Modeling, Extract, Transform and Load, an adaptation of the traditional ETL process including the Modeling phase as a prior requirement. As an implementation use case, laboratory test results of a population of 230000 patients belonging to Troms and Finnmark counties in Norway requested between 2013 and 2014 have been standardized. Tests registers were

normalized by defining transformation and aggregation functions between them and a previously defined archetype. These mappings were used to automatically generate openEHR compliant data. These data were loaded into an archetype-based data warehouse. Once loaded, we defined indicators linked to the data in the warehouse to monitor test activity of Salmonella and Pertussis using the Archetype Query Language. The execution of the project proved that this archetype-based approach simplifies the process of reusing clinical data for clinical research, clinical decision support and public health surveillance. In turn, this means an improvement of health care processes in North Norway Region, specifically in the detection of Pertussis and Salmonella outbreaks, whose incidence has suffered a significant increase since 1997 in that region.

- **Related publication:** L. Marco-Ruiz, D. Moner, J.A. Maldonado, N. Kolstrup, J.G. Bellika, Archetype-based data warehouse environment to enable the reuse of electronic health record data, *International Journal of Medical Informatics*. 84 (2015) 702–714. doi:10.1016/j.ijmedinf.2015.05.016.

1.3.4. Grants

The author was beneficiary of grant DI-14-06564 (*Doctorados Industriales*) of the *Ministerio de Economía y Competitividad* of Spain, between 2016 and 2018. The objective of this grant is to support the development of PhD thesis inside private companies. Thus, the research results are grounded in real needs of the market and provide a direct benefit to the company activity. In this case, the thesis was partially developed in VeraTech for Health, a spin-off company of the Technical University of Valencia. The author is also co-founder of this company.

The author is beneficiary of grant INNTA3/2020/12 (INNODOCTO) of the *Agencia Valenciana de la Innovación, ayudas del Programa de Promoción del Talento – Doctorados empresariales*. The objective of this grant is to promote the acquisition of skills to develop research and development projects in business environments. Part of the research of this thesis was developed in VeraTech for Health during the year 2020.

1.3.5. Projects

The author of the thesis has participated in 16 R&D projects in the field of semantic interoperability of health information. These projects served to learn about the problems and needs for creating and managing clinical information models, and to validate and improve the thesis results. The main projects are listed next, and they are further described in Chapter 6.

- **Spanish National EHR project (HCDSNS).** The aim of this project is to develop an infrastructure that allows sharing relevant and interoperable health data sets (documents) between the seventeen regional authorities in Spain. Forty-six professionals from 27 scientific societies, and 21 institutional representatives, participated in the definition of contents of the exchanged documents. Based on those definitions, we implemented 64 ISO 13606 archetypes and 36 HL7 CDA archetypes. The first release of the archetypes only covered the structural definition of the CIM, i.e. they define only the information structures according to the selected RM. In a second phase, the Ministry of Health defined a SNOMED CT national extension to support all the new concepts used in the archetypes, and the appropriate reference sets for data entry points (approximately 70 reference sets). In this project, we also developed an archetype publication and governance system (*Servidor de Objetos de Modelado, SOM*).
- **Uruguay National EHR project (HCEN).** The aim of the project is to build archetypes and implementation specifications of six clinical documents for the National Unified EHR system. A group of clinical experts studied, with the support of the technical experts, existing literature and standards, and defined the contents of the archetypes corresponding to the six clinical documents. Technical experts implemented the specifications into formal and standardized resources (41 ISO 13606 archetypes and 6 HL7 CDA implementation guides). The terminology expert defined SNOMED CT subsets to populate the possible values of information elements such as procedures, allergies and diagnosis. Finally, all the generated materials and specifications were published in an online management system.

- **Quality-assessed maternal-child care data repository.** In this project, we developed an integrated, standardized, and quality assessed, data repository for research and monitoring of best practices of mothers and newborns care. We defined the contents of the repository in the form of ISO 13606 archetypes. Afterwards, we defined 112 clinical indicators about the clinical practice on newborns were defined, based on the archetype data structures. Finally, we normalized and incorporated into the repository information from two different hospitals (Hospital Virgen del Castillo in Yecla, and Hospital 12 de octubre in Madrid) and more than 5700 newborns.

1.3.6. Software

The LinkEHR Interoperability Platform (www.linkehr.com) is a set of software modules that supports and facilitates the integration and semantic interoperability of health information by means of EHR standards and archetypes.

As part of the platform, the author of this thesis has particularly contributed to the development of two modules:

- **LinkEHR Studio.** A reference model-independent archetype editor and data transformation program. LinkEHR Studio incorporates all kind of functionalities to facilitate archetype development, including different representations of the archetypes to support the different stages of development.
- **LinkEHR Model Manager.** A clinical information model management system. It is designed to facilitate the governance of semantic interoperability artefacts such as archetypes, and other implementation resources such as implementation guides, instance examples, or domain documentation.

1.3.7. Other contributions

During the development of the thesis, the author has been in charge of many formative actions covering topics such as semantic interoperability principles, ISO 13606, openEHR, HL7 CDA, HL7 FHIR, SNOMED CT, clinical information and arche-

type modeling, and data quality. In total, more than 300 people have received training in those standards and technologies. The author was responsible of the development of the training materials, the preparation of practical exercises, and responsible of giving courses that ranged from half a day to three-day courses. In particular, the author was the main contributor to the creation of a 15-hour on-line training course about ISO 13606 for the Spanish Ministry of Health.

Additionally, between 2015 and 2018, the author has actively participated in the development of the ISO 13606:2019 norm in the ISO/TC 215 committee of health informatics. In this period, he has been in charge of updating the data types of the reference model, and he has collaborated to the revision of the complete reference model of this norm.

1.4. Structure of the thesis

Table 1 presents the overall structure of the thesis. Chapter 1 introduces the main hypothesis of the thesis, together with the research questions, research objectives, and contribution of the thesis. Chapter 2 describes the basic elements of semantic interoperability, namely electronic health record standards, archetypes, and medical terminologies. Chapter 3 develops a study of existing methodologies and systems that provides the basic elements to define an archetype modeling methodology and the basic principles of archetype governance. Chapter 4 introduces the archetype modeling methodology, a formal proposal for creating quality archetypes. Chapter 5 identifies and defines the principles and characteristics of archetype governance. Chapter 6 describes the tools that the author, in collaboration with his research colleagues, has developed to support archetype development and governance. In addition, it provides a description of some research and development projects where the results of this thesis have been used and improved during the last decade. Finally, Chapter 7 summarizes the contributions of the thesis and introduces possible future works.

Chapter	Contents	Research questions	Objectives	Contributions	Publications
1	Motivation of the thesis, hypothesis, and main contributions				
2	Introduction to the elements of semantic interoperability		O1		
3	Study of archetype modeling methodologies and governance initiatives	R1, R3	O1	C1, C3	P1
4	Description of the archetype modeling methodology	R1, R2	O2	C2	P1, P2, P3, P5, P6, P7, P8
5	Description of archetype governance principles	R3	O3	C3	P4, P8, P10
6	Software and applications of the results of the thesis	R4	O4	C4, C5	P2, P5, P6, P7, P9, P11, P12
7	Conclusions and future work				P1, P4

Table 1. Structure of the thesis and relationship with the research questions, objectives and contributions

Chapter 2

Electronic health record formalization and semantic interoperability

Health care produces and consumes large amounts of information, which is the basis of accurate and efficient health care. Any clinical decision or act must be based on information about the patient, not only on what can be observed at the present, but also include retrospective and even prospective information. All this information should be available, not only during health care processes but also to support clinical research or any other activity that makes use of existing health data.

This chapter introduces the basic concepts of health information systems, and the electronic health record as the core repository of health data. It also describes the concept of semantic interoperability of health data and its core components: EHR standards, clinical information models and clinical terminologies.

2.1. Health information systems and electronic health records

The management of health information relies in the use of specialized information systems applied to the medical domain (Figure 1). An information system is a provision of information, processes, people and information technologies that interact to collect, process, store and provide as a result the information necessary to support the organization [2]. In the health sector, information systems include a wide range of applications, from data entry systems to decision support systems; used by multiple types of users, such as patients, health professionals, IT profes-

sionals, employers, insurance companies, pharmaceutical firms, health provider organizations, and governments. Thus, we can adapt the previous definition of information system applied to the health domain: a health information system is a provision of information, processes, people and information technologies that interact to collect, store, process and provide as a result the necessary information for health care.

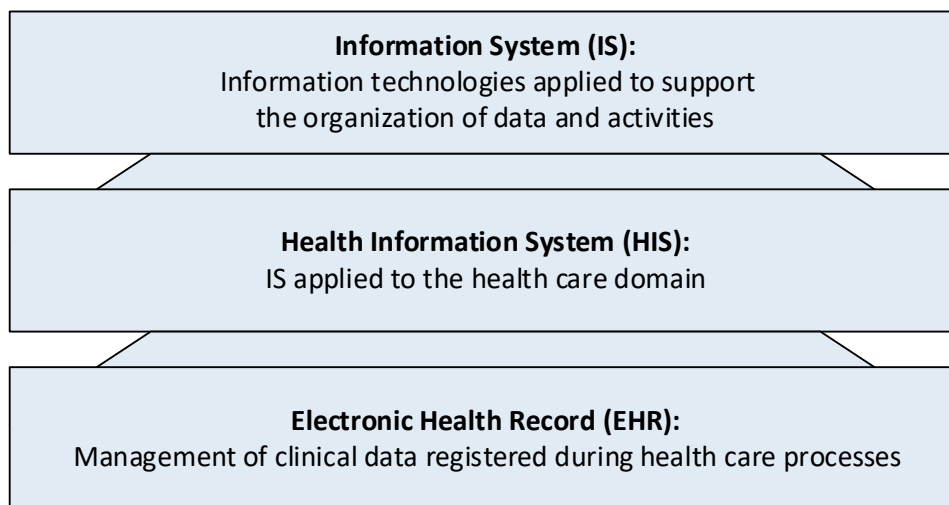


Figure 1. Specializations of information systems in the health domain

Health information systems have some particular characteristics. The first characteristic is the diversity information handled. Diversity includes aspects such as professional disciplines, treatment options or healthcare procedures. Another important aspect is that they directly affect the health status of people. Medical professionals need to have all the relevant information about the patient's health in order to take the correct decisions during the provision of care. Otherwise, medical errors with serious consequences, including the death of the patient, can occur. Information systems should guarantee and maximize the quality of the information they manage and that is provided to health professionals. Another important characteristic of health information systems is that the management of health information is highly related to privacy and confidentiality issues. Any use or exchange

of information involves a risk that the information may be accessed by non-authorized parties. There is a need of balancing health information availability and health information privacy, but the second should not limit the provision of meaningful information to health professionals during healthcare.

Electronic health record systems (EHRs) are the core of health information systems. Their purpose is to provide support to all the information collected and required during health care for patients. This includes information to support primary health care, information from particular specialties or services (e.g. the pathology laboratory, radiology service, operating theater, admission, etc.), information to support remote health care through video, voice or text messages (telemedicine), or information to support decision support systems, which help in the decision-making process of health professionals, just to mention a few.

There are numerous definitions describing the scope of an EHR, specifying the information they should contain, or the functionality they should offer. The International Organization for Standardization (ISO) recognizes this difficulty: “Previous attempts to provide a definition of the EHR have failed due to the difficulty to cover each and every one of the aspects of the EHR in a single and complete definition” [3]. This also provides an indication of the complexity of formalizing the contents of an EHR. However, in the same report, ISO provides a definition of an EHR: “Repository of information on the health status of a subject of assistance, in a computable format, stored and transmitted in a secure and accessible by multiple authorized users. It has a standardized or consensual logical information model, which is independent of health care support information systems. Its main purpose is to provide continuous, quality and efficient health care, which contains information that is retrospective, current and prospective.”

The introduction to the ISO 13606 standard provides another definition [4]: “The EHR is the persistent, longitudinal, and potentially multi-enterprise or multinational record of health and care provision relating to a single subject of care (the patient), created and stored in one or more physical systems in order to inform the

subject's future health care and to provide a medico-legal record of care that has been provided."

From both definitions, we conclude that the EHR does not belong to a single health care provider, but it can potentially be distributed in multiple places; that it contains a record of the care provided to a person (patient) during his whole life; and that its main purpose is to facilitate and to document health care.

EHRs have additional characteristics that are widely recognized [5]:

- **Availability.** All authorized personnel can simultaneously access information about patients at the time they need it, not only locally but also remotely.
- **Readability.** The information is generally more readable than in paper format (manual writing) and it is better organized because the computer system imposes a structure on the data entered.
- **Completeness and consistency.** The use of computers makes it possible to improve the completeness of the data by being able to define default values and force the completion of some fields. On the other hand, it allows the automatic application of checks on the data as they are registered in the system. For example, for the validation of reference ranges for numerical results.
- **Flexibility.** Allows the presentation of data in various ways: chronologically, by problems, by data source, etc.
- **Reuse.** Facilitates medical research and teaching. Potentially, it allows locating those relevant clinical histories given a certain pathology, treatment, social situation of the patients, etc. It also allows the implementation of repositories or clinical data banks such as disease registers.

All the previous characteristics are based on simple principles: it is necessary to have the information well defined and structured, and it is necessary to be able to exchange health information seamlessly between different systems. These principles are also the foundation of semantic interoperability.

2.2. Semantic interoperability of health data

The implementation of EHR systems is a reality worldwide, especially in developed countries. According to the Healthcare Information and Management Systems Society (HIMSS), 78% of European health organizations manage digital patient data [6]. In the United States, the Office of the National Coordinator for Health Information Technology (ONC) totals an 85.9% of adoption of EHRs for Office-based physicians [7]. In contrast, many other countries still have a low adoption of EHRs, although it is recognized their need to improve healthcare [8].

Increasing the adoption of EHRs is necessary but not sufficient to fulfil all the EHR characteristics mentioned in the previous section. It is also necessary that existing EHR systems can communicate and to understand the communicated data automatically. This is not a problem unique to the health domain. In other sectors, there are solutions that guarantee the ability of the systems to communicate with each other. Some examples are mobile phone technologies, financial services or travel booking agencies. In all these cases, agreements have been reached and standards have been developed so that services and information from different sources can be used together.

In the case of health information, the ultimate goal is to make it accessible from anywhere and at any time, to be used transparently and automatically by any information system and, therefore, usable by health professionals during health care or for information reuse.

Interoperability is the ability of computer systems to work together (interoperate), through the exchange and use of the information handled by each one of them. Existing literature distinguish four types of interoperability [9]:

- **Foundational interoperability.** It establishes the inter-connectivity requirements needed for one system or application to share data with and receive data from another.

- **Structural interoperability.** It defines the syntax of the data exchange. It ensures that data exchanges between information technology systems can be interpreted at the data field level.
- **Semantic interoperability.** It is the ability of two or more systems to exchange data and information, maintaining their original meaning and being able to receive, interpret, aggregate, analyze and work with the received information as if it were their own automatically.
- **Organizational interoperability.** It refers to the ability of organizations to communicate and interpret information between them despite using different systems, protocols and policies, or meeting in different cultures.

Current technologies can solve most of the foundational and structural interoperability difficulties, but semantic interoperability is still a focus of current research and development efforts [10]. The complexity of semantic interoperability arises from the need of making health information meaningful and automatically computable by external information systems. To that end, the following requirements are needed.

- To provide a safe and reliable environment for the communication and aggregation of clinical information between heterogeneous systems.
- To link and align the information of the EHR with standard clinical terminologies, as well as other knowledge resources that may be available (clinical information models).
- To link the information of the EHR with contextual, explanatory information or with educational materials that helps the patient and their relatives to understand clinical information and health professionals' comments.
- To guarantee the quality and consistency of exchanged data to support data reuse for public health, research, management of health services, etc., always under the premise of respect for the safety and privacy of the patients.

Semantic interoperability will only be feasible if there are agreements about standards, information models, terminologies and semantic definitions used for the shared data. We must consider that social, cultural, and legal aspects within each

organization, region or country will influence the implementation of semantically interoperable systems. It will be necessary that data providers and data consumers think beyond their own needs and constraints and accept a minimum set of technical elements that allows them to work together. Semantic interoperability relies on three basic elements [1], which will be described in detail later.

- The use of a generic **reference model or standard** to represent the data homogeneously, safely, and completely.
- The use of **clinical information models** that define structures of information with a specific meaning or for a specific use.
- The use of shared **clinical terminologies** that define the vocabulary used to describe the data.

2.3. Electronic health record standards

EHR standards provide a framework to represent clinical data in a common and generic format. During many years, EHR systems followed proprietary designs to satisfy a set of functional characteristics defined by specific users or clients. The adaptation or extension of those systems to be used in different contexts was limited. It was even difficult to adapt those systems to satisfy new requirements of the same users. Another problem was that EHR systems tend to include only the minimum set of information required to support clinical care. The EHR did not register many information not directly related to the patient health status. For example, information about the context of the health care encounters, or to the evaluation made by health professionals. This limits the reuse and future interpretation of health data, or at least, makes that future interpretation potentially unsafe.

To mitigate these problems, the latest EHR standards were developed following two key principles: to be flexible when representing clinical data structures, and to allow the representation of the context information that accompanies that clinical data.

The first principle implies that the data structures of the EHR are not fixed. They can be adapted to support new information and use cases. To achieve this, new EHR standards define generic data models loosely coupled to specific concepts and information structures of the health domain. They only represent the common and minimal characteristics of clinical data. Most of the complexity of the health domain data is represented by complementary clinical information models, described in Section 2.4.

The second principle is to support a complete representation of context information. We define context information as any data that informs about how data was registered, or that influences in the interpretation of clinical data contained in the EHR. For example, a diagnosis of diabetes mellitus within a "family history" section does not imply that the subject of the information suffers from this disease. EHRs are also a legal record of the health activity. For example, an EHR has to record the participants and location of a clinical act for future reference. Therefore, the clinical data must be accompanied by the necessary information required by current legislation, and to support audit or control tasks.

Among the health information standards and technical specifications that fulfil the previous characteristics, we can highlight the following ones [11]:

- **openEHR** [12]. Open specifications for the creation of EHR systems, including the storage, retrieval, and exchange of health data. OpenEHR defines a Reference Model (RM) that includes the basic characteristics and structure of clinical data, independently of any particular technology of implementation. OpenEHR RM defines the following building block for any EHR: FOLDER, COMPOSITION, SECTION, OBSERVATION, EVALUATION, INSTRUCTION, ACTION, ITEM_TREE, ITEM_LIST, ITEM_TABLE, and ELEMENT. The ELEMENT is the container of clinical data represented using a specific DATA_TYPE. The RM also represents the basic context information needed to safely interpret and reuse registered clinical data.

- **ISO 13606** [4,13]. Standard firstly released in 2008 and renewed in 2019 for the communication and semantic interoperability of EHR extracts. An EHR extract is a partial or complete view of an EHR of one person that is collected for its communication between different EHR systems. ISO 13606 defines a RM, similar to the one of openEHR, but using even simpler building blocks: FOLDER, COMPOSITION, SECTION, ENTRY, CLUSTER, and ELEMENT (Figure 2). As in openEHR, context information always accompanies these classes.
- **HL7 CDA** [14]. Standard for the representation, persistence, and communication of clinical documents. It is both a human readable and machine computable representation of clinical data generated during clinical encounters, or because of a documentation activity. HL7 CDA documents are built upon a generic Reference Information Model (RIM), which includes a header with metadata information, and a body with the clinical data. The body includes a readable representation of the data, but it can be also extended to structure data into sections and structured clinical statements (Act, Observation, Substance administration, Encounter, Procedure, etc.).
- **HL7 FHIR** [15]. Standard framework for the creation of a web service infrastructure for the communication of clinical information. HL7 FHIR focuses on implementation, by providing the definition of modular portions of clinical information (Resources) for fast and easy implementation using web standards, for example, using REST web services as the technological infrastructure for the exchange of clinical data.

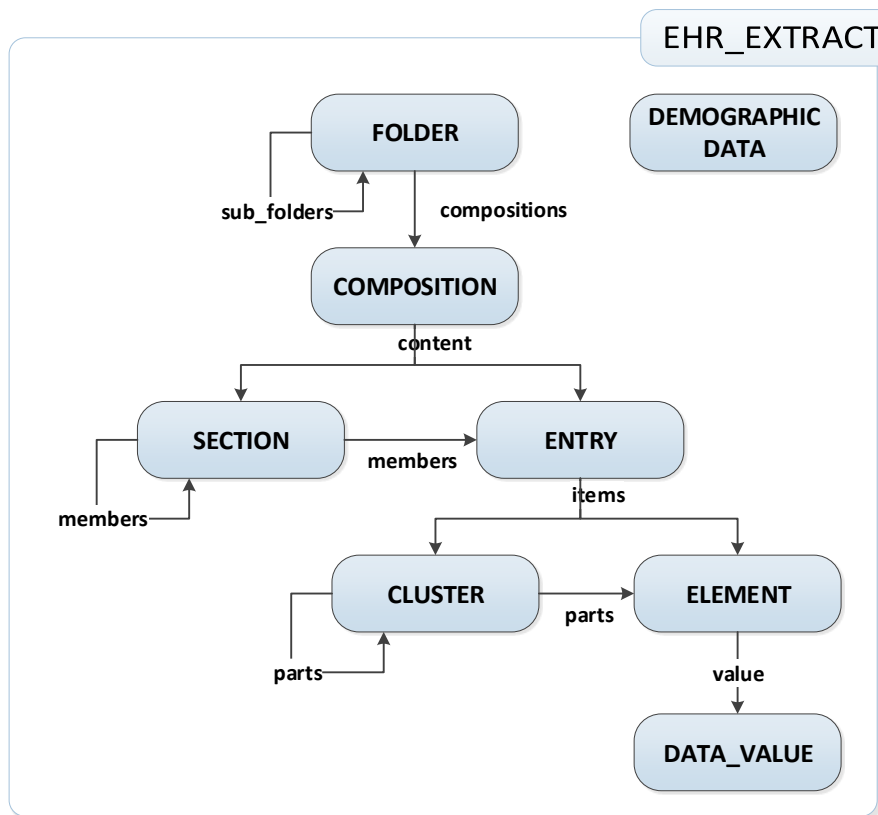


Figure 2. Example of a standard reference model: main classes of ISO 13606

Not all the mentioned standards have the scope of representing a complete EHR. That is only the case of openEHR. ISO 13606 can also represent all the information of an EHR, but it lacks a proper support for other relevant modules such as the representation of local information templates, or a query language for the EHR. The other standards only represent partial views of the information contained in the EHR. However, they all follow a similar architecture: they provide a generic information model, and they use clinical information models to define meaningful data structures.

2.4. Clinical information models

In this thesis we use the expression clinical information model (CIM) as a generic term that encompasses all technical specifications defining how clinical information is organized and described inside an EHR system, a clinical data repository or for EHR communication. A CIM defines both the information structure and formal semantics of clinical data. CIM are structural and semantic artifacts that facilitate organizing, storing, querying and displaying clinical data, exchanging that data between different information systems, and performing data analytics. Usually, a CIM is defined by constraining the generic data structures of an underlying reference model that provides the basic characteristics and attributes needed to represent data instances.

The author of this thesis contributed to a systematic review of clinical information modeling processes that was published in [11]. In this section, we include an excerpt of that work. *Goossen et al.* also described initiatives that follow a CIM approach, indicating their differences and similarities in [16].

The HL7 v3 modeling approach is based on a standard Reference Information Model (RIM) representing the main business logic of any healthcare environment, from which specific messages and documents can be defined. HL7 v3 messages [17] and HL7 CDA are standards based on the HL7 RIM. It is possible to define clinical information models for HL7 CDA in the form of HL7 templates that specify how to organize the clinical information within each kind of document, for specific clinical communication purposes.

HL7 FHIR uses modular components called Resources. These resources (definitions of common reusable patterns of clinical information) can be combined or extended in order to provide particular solutions to health information systems. Therefore, they are also a kind of CIM.

Another important modeling approach is based on the dual level methodology [18], based on the definition of a synthesized and generic Reference Model (RM) that is designed to represent the most basic properties and structures of any

EHR. Clinical information models are defined in the form of archetypes. Archetypes define how data have to be structured in order to be seamlessly stored or transferred between EHR systems. The dual model approach is supported by the EN ISO 13606 standard and the openEHR specifications.

Additional modeling approaches have emerged focused on defining generic information models at a conceptual level, without depending on a specific implementation. The Clinical Information Modeling Initiative (CIMI) [19], Detailed Clinical Models (DCM) [20], and the Clinical Element Model (CEM) initiative [21] are examples of such generic models.

Figure 3 summarizes the reference models used (i.e. the models that represent data instances) and the CIM technology employed by each of these initiatives.

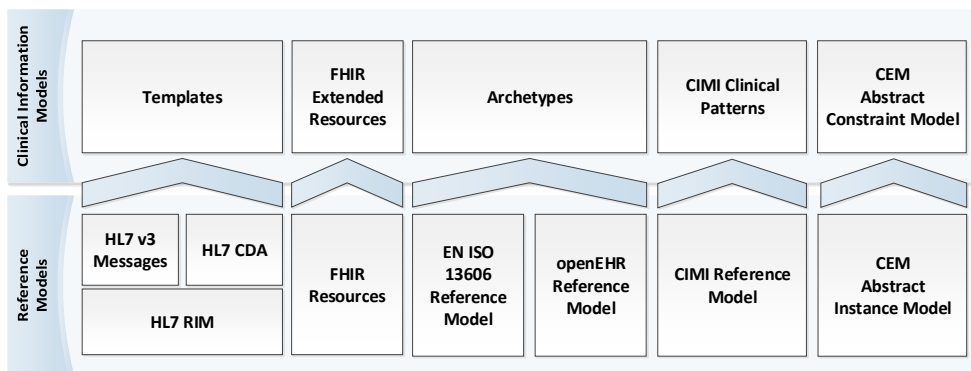


Figure 3. Summary of Reference Models and their Clinical Information Model definition artifacts, extracted from [11]

Although all CIM approaches have similar characteristics, this thesis will focus on archetypes as the main candidates for representing CIMs. Archetypes are formal definitions of domain-level concepts, in the form of structured and constrained combinations of the classes and possible instance-values of a RM. Their principal purpose is to provide a powerful, reusable and interoperable mechanism for managing the creation, description, validation and query of EHR data. Examples of structures of information in the form of archetypes include prescriptions, problem lists, differential diagnosis, pregnancy reports or blood pressure observations. In fact, any desired set of information of the EHR can be defined as an archetype.

Technically, an archetype definition is a tree representing an Object-Oriented model, i.e. it is composed by a nested succession of classes and attributes. An archetype definition process consists in creating this structure for a specific use case, including constraining the data values that are acceptable or valid (Figure 4).

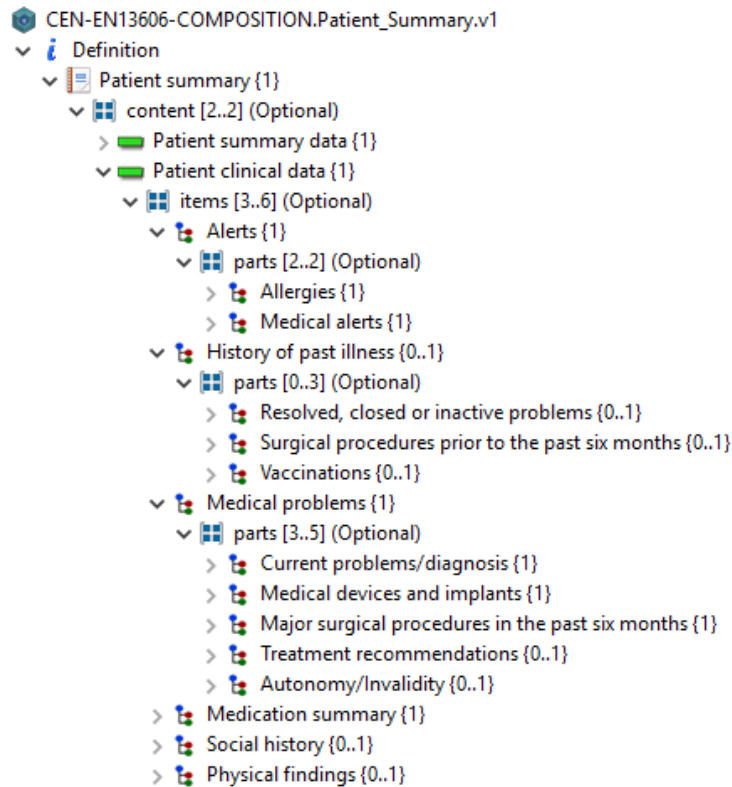


Figure 4. Example of a patient summary archetype

One interesting property of archetypes is their reusability. We can reuse archetypes by aggregating them into a bigger one. Thus, we can construct complex archetypes based on simpler ones without repeating the data structure. We can also reuse an archetype by specialization. This is achieved by providing further constraint the structure already expressed by other archetype. Since this process can be repeated as many times as needed, there exists an underlying specialization hierarchy behind every archetype whose root is a RM class which has a succession of children archetypes. The deeper the level of the hierarchy is, the more constrained or specialized the archetype is.

However, archetypes are not only a structural definition of data. They are also the link between information structures and their semantic description. Once an archetype structure is created, terminology bindings can be specified to describe the meaning of that structure. Moreover, terminologies can also be used to describe the set of data values or valid codes that can be used as data items during the clinical data instance generation. Therefore, clinical terminologies are the third basic element of semantic interoperability.

2.5. Clinical terminologies

Clinical terminologies are an essential part of the documentation of health care. As defined by Alan Rector, “clinical terminology concerns the meaning, expression, and use of concepts in statements in the medical record or other clinical information system” [22]. They provide a coded and controlled vocabulary that identifies univocally the different concepts involved in medical science.

Health professionals use clinical terminologies as controlled vocabularies, to annotate portions of the health data of patients recorded during a health care activity. The intention is to mitigate the vagueness of natural language, thus improving the accuracy of health records.

A meaningful analysis of clinical recorded data also requires the use of terminologies. Examples of these analyses include the classification of diseases to study their prevalence, the analysis of costs of health care, and the support of population health or facilitating clinical research. Not using clinical terminologies to describe formally the registered clinical data would hinder all these activities.

There is a broad variety of clinical terminologies. Their coverage ranges from very specific domains, to comprehensive terminologies that nearly cover the complete health domain. They can be highly granular and define very specific concepts, or define only generic concepts for classification or aggregation purposes. Any of these approaches can be useful depending on the scenario of use.

Some examples of widely used terminologies are:

- **SNOMED CT.** A global terminology for healthcare. It allows the representation of nearly any clinical information in a coded form, with a high level of expressivity. SNOMED CT not only provides a set of coded concepts and terms, but also the relationships between concepts. This provides support for meaning-based retrieval, analysis and reporting of clinical data [23].
- **ICD-10.** The 10th revision of the International Statistical Classification of Diseases and Related Health Problems is a medical classification published by the World Health Organization (WHO). It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases [24].
- **LOINC.** Logical Observation Identifiers Names and Codes is an international terminology for identifying health measurements, observations, and documents [25]. It is widely used in clinical laboratories reporting, but its use is increasingly spreading to other areas.

Clinical terminologies not only describe health data instances, but also play an important role in the initial definition of archetypes. Archetypes are both a definition of information structures, and a concept definition, that is, archetypes are collections of information items with a specific meaning. Clinical terminologies are the mechanism to describe that meaning.

Once we have modeled an information structure in the form of an archetype, we can bind it to clinical terminologies to provide a univocal definition of the model. We can bind archetypes and terminologies in two different ways:

- **Value binding.** The primary use of terminologies is to annotate data. Since one of the objectives of archetypes is to define a schema for data, they also must guide how we can register coded data. First, we specify value sets from a terminology, i.e. sets of possible codes that can be used to annotate a type of coded data. For example, a value set for oncology diagnosis, or a value set for pediatric procedures. There are two strategies to create a value set: using an extensional definition or using an intensional definition. An extensional definition of a coded value set means that we explicitly and individually select the concepts

that are part of the value set. An intensional definition of a value set means that we specify the properties that the selected codes must fulfill in the form of an executable expression or query to the terminology server. As an example, SNOMED CT provides the Expression Constraint Language [26] to define intensional value sets. Second, we have to assign a value set to coded data elements of the archetype. This means to constrain the possible values that coded elements of the archetype can register to the values included in a specific value set.

- **Semantic binding.** We can use terminologies to specify the meaning of the archetype elements. An archetype is a definition of a clinical information structure. Although it follows the rules of an underlying reference model, it is not enough to provide a specific meaning of those information structures and data elements. We can use terminologies to describe the meaning of each of the archetype data elements. This will allow the automatic and semantic processing of archetypes, and thus, it is an essential part to make archetypes semantically interoperable.

For example, a blood pressure archetype specifies the data elements that can be registered in the EHR (i.e. systolic and diastolic values, both measured in mmHg, and greater than or equal to zero). At the same time, the blood pressure archetype incorporates the semantic mapping of the data elements by adding mappings to medical terminologies (i.e. the archetype structure is mapped to SNOMED CT concepts 75367002 |Blood pressure (observable entity)|, 271650006 |Diastolic blood pressure (observable entity)|, and 271649006 |Systolic blood pressure (observable entity)|). Finally, according to the registered data, a hypertension diagnosis might be recorded in the EHR as 38341003 |Hypertensive disorder, systemic arterial (disorder)|

Clinical terminologies are large. ICD 10 has nearly 72,000 diagnostic codes, LOINC has more than 91,000 codes, and SNOMED CT has more than 340,000 concept codes, nearly 900,000 terms and 2,750,000 relationships. It becomes neces-

sary to use the appropriate software to manage, explore and use clinical terminologies. This is the role of terminology servers. Terminology servers are specialized systems that provide an accessible repository of medical conceptual knowledge (concepts), potentially mapped to different natural languages (terms) [27]. Terminology servers are usually accompanied by a terminology browser that facilitates the exploration of the terminologies. The functionalities provided by terminology servers can include searching codes to obtain their textual description, searching natural text to obtain candidate codes, search for translations of a code into different languages, search for mappings of a code to different terminologies, the management of versions of the terminologies, or the creation, management and publication of value sets. These functionalities can be offered using a visual user interface or accessed through a service API, such as CTS2 [28] or HL7 FHIR.

We can use these APIs to connect a terminology server to other systems, such as archetype editors, to support terminology mappings during the archetype modeling. However, the incorporation of terminologies to archetypes needs a methodological solution in addition to the technical solution. The creation of archetypes must follow a clear strategy, a formal modeling process that guides both the definition of the information structures and the definition terminology mappings. This will help in creating quality archetypes where the information structures are aligned with terminologies. Moreover, once the archetypes and value sets are created, we will need a governance process to maintain them up to date to new requirements. We introduce these aspects in the following sections.

Chapter 3

Study of archetype modeling methodologies and governance initiatives

Two of the components of semantic interoperability mentioned in the previous chapter already have mature methodologies guiding their development. EHR reference standards, when implemented as software, follow existing software engineering and logical database design methodologies [29–35]. There are also several development methodologies for creating terminologies and ontologies [36–40]. However, CIMS, and in particular archetypes, lack a well established methodology for their development, as concluded in a previous study [11]. They also lack governance guidelines that define the context where archetypes are developed, maintained, and published, i.e. archetype governance guidelines.

To develop a proposal for an archetype development methodology and the guidelines for their governance, it is necessary to study and analyze existing experiences in this area. In this chapter we provide a comparative study of experiences documented in the literature to learn the requirements of archetype modeling methodologies and archetype governance principles.

3.1. Modeling of archetypes

Clinical experts are usually in charge of developing archetypes in the context of a local, regional, or national EHR project, where archetypes provide a formal description of the clinical information to be used or shared. The benefits of using an archetype modeling methodology are twofold. Firstly, the methodology helps in the coordination of the development team and provides a set of tools and strategies to ease and accelerate the development process. Secondly, it facilitates the reuse of

archetypes created by different teams. Following ad hoc methodologies, archetype authors might produce archetypes that are only usable in their own projects, but not in other contexts. New archetypes might overlap or be incompatible with existing ones. This is the result of considering archetype creation as a craft rather than an engineering process. To mitigate this problem, we need to provide archetype authors with specific and formal rules that guide archetype development and governance. A clear archetype modeling methodology is necessary to define reusable and sound archetypes.

A characteristic of archetypes is that they can combine, in a single artefact, the specification of information structures, and their semantic description through medical terminologies. Due to this close relationship, and to achieve meaningful results when building archetypes, archetype modeling methodologies should be aligned to the modeling of software, terminologies and ontologies.

Table 2 provides a short summary of methodologies that influence the design of an archetype modeling methodology. A more detailed description of the activities covered by each type of methodology is included in the following sections.

Methodology domain	Resulting artefact	Scope of the modeled artefact	Example artefacts	Example methodologies
Software modeling	Data model	Model of the basic data structures of the EHR	ISO 13606 Reference Model OpenEHR Reference Model	Software engineering methodologies, data base methodologies, IEEE 1074 standard, ISO/IEC 12207 standard...
CIM modeling	Clinical Information model	Model of the domain-oriented information structures	ISO 13606 archetypes OpenEHR archetypes	Only partial experiences are found in the literature without describing the development methodology in detail
Ontology modeling	Concept model	Model of meaning including medical vocabularies, concept definitions and their relationships	Open Biological and Bio-medical Ontology (OBO)	UPON, On-To-Knowledge, TOVE, IDEF5, METHONTOLOGY...

Table 2. Summary of modeling methodologies and their domain of use

3.1.1. Software development methodologies

Software development methodologies and software engineering provide a reliable reference for developing archetypes from a technical perspective. Archetypes are part of the software specifications of EHR systems since they constrain a particular object-oriented RM, i.e. a model that can be implemented in systems to store and communicate data instances. Thus, there is a direct relationship between archetypes and technical implementations of information systems, which in turns implies a relationship between archetype modeling and software development.

There are many software development methodologies. For example, the traditional waterfall model [29], the incremental build model [30], the spiral model [31], or the more recent agile software development methodologies [32]. Beyond their specific differences, they share common development phases: requirements analysis, functional and technical design, implementation, testing and installation. Several standards have formalized the process of software development. One of them is the IEEE 1074 standard for developing a software project life cycle process [33]. It provides a basic common framework to be followed by any specific software development methodology. Similarly, the ISO/IEC 12207 standard [34] establishes guidance for software life cycle processes, activities and tasks during the development, operation, maintenance and disposal of software products.

It is also worth mentioning logical database modeling [35], whose objective is to define the most optimal and detailed data model of a database. This includes identifying entities, their attributes, relationships between the entities, and normalizing the data model. These activities are also relevant for archetype modeling.

3.1.2. Clinical information modeling methodologies

In a previous work [11] the author contributed to a systematic review of CIM development methodologies found in over 50 published papers. The conclusion of the systematic review was that a unified modeling methodology does not exist. Most of the existing CIM developments shared a similar approach, although only in

few cases the authors clearly described the methodology followed. The identified common phases included the scope definition, the analysis of the information covered in the specific domain, the design of CIMS, the definition of implementable specifications, the validation of CIMS, and their publication and maintenance.

In the case of archetype modeling, the most relevant reference is the openEHR data modeling approach (ODMA) [41], a five step methodology used by some archetype authors. However, only a sparse description of the methodology steps is found in the literature. There is no detailed information about the complete methodology workflow, the expected inputs and outputs of each step, or about the required participants during the modeling process. Moreover, ODMA is focused to the development of openEHR archetypes and its tooling ecosystem, and not a generic archetype modeling methodology. All these factors limit the applicability of ODMA.

3.1.3. Ontological modeling methodologies

There are several examples of ontology development methodologies. A non-exhaustive list would include the United Process for Ontologies (UPON) [36], On-To-Knowledge [37], TOVE project [38], IDEF5 [39], or METHONTOLOGY [40].

The objective of these methodologies is to provide guidelines about the specification, conceptualization, formalization and implementation of ontologies. They share the following phases:

1. Specification of the scenario and scope, mentioned in all the analyzed methodologies [36–40]. This phase establishes the purpose, context, and scenarios of use of the ontology to be developed.
2. Knowledge acquisition, mentioned in [36,37,39,40]. It includes the collection of relevant data, documentation and existing ontologies in order to analyze, and refine the ontology requirements.

3. Conceptualization. It consists on the structuration of the domain knowledge in terms of concepts, attributes and relationships. It also includes the development of a terminology or glossary of terms for the ontology. All the analyzed ontology methodologies include this phase.
4. Implementation. In this phase, the ontology is encoded using a formal language such as OWL. It requires the use of a development environment with editors, syntactic analyzers, and validators that help in detecting inconsistencies or redundant knowledge. Again, all the studied methodologies include this phase.
5. Evaluation, mentioned in [36,37,39,40]. The evaluation phase guarantees that the ontology is fit for purpose, and that it meets all the initial requirements. It may also evaluate the syntactic and semantic quality of the ontology.

In addition to these activities, only one methodology [37] mentions the application of the ontologies and their future evolution and maintenance. It is also the only one to mention the tools and participants in the ontology development process.

3.2. Governance of archetypes

Archetype development usually takes place in the context of a broader archetype governance process that identifies the need of developing a new set of archetypes, and controls how to use them after their development.

The concept of archetype governance derives from the existing concept of IT governance [42,43]. IT governance provides the technological infrastructure and infostructure needed to support a business strategy. In the context of health information systems based on an archetype approach, archetype governance focuses in the infostructure aspect, since archetypes are the mechanism for the definition of the data structures of health information systems. Archetype governance supports

the process of creating and managing archetypes and their relations among them and to other semantic interoperability resources such as terminologies.

Health information systems designers rely on archetype governance to coordinate archetype development, to identify the need of new archetypes, to identify the obsolescence of existing archetypes, or to avoid overlap between archetypes. Despite its importance in building health information systems, there are few references to archetype governance in the literature. Moreover, they are inconsistent in defining the scope of the governance process. It becomes necessary to identify the different areas of knowledge governance to understand where to locate archetype governance (Figure 5).

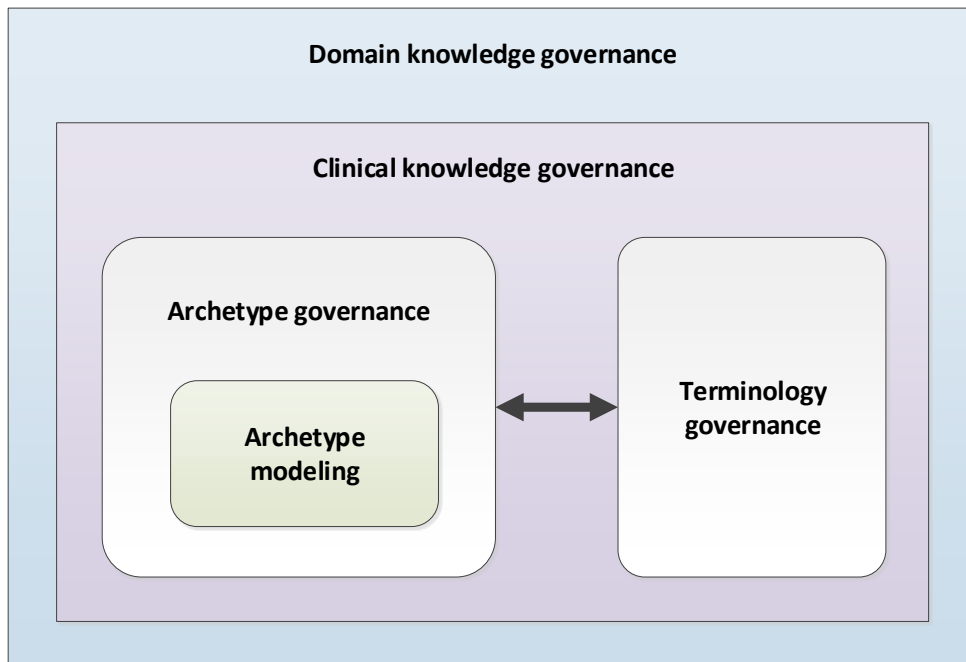


Figure 5. Areas of knowledge governance

Domain knowledge governance, or knowledge management, is “all tasks related to establishing (...) formal and informal organizational mechanisms and structures in order to systematically influence the building, dissemination and maintaining of knowledge within and between domains” [44]. It is a generic denomination

applicable to any organization that relies in information technologies, information structures and any other knowledge resources as part of their business activities.

In the case of the clinical domain, some authors introduce the concept of clinical knowledge governance (CKG) [45,46]. CKG covers the management of all kinds of knowledge resources used in a clinical domain, including archetypes, templates (localized archetypes), and terminology value sets. These resources are essential for the design, development and use of clinical information systems. As it has been previously explained, the definition of any archetype information structure must be completed with terminology bindings, to make the archetype semantically interoperable. The management of archetypes and terminologies is closely related, but they require specific rules of governance. For example, during the governance of archetypes it is important to maximize the reuse of existing information structures. It is also important to adjust the data structure to the specific local needs of the users, and to provide a mechanism to modify archetypes and make them publicly. In the case of terminologies, it is important to define the adequate value sets containing the codes needed at each place of the information structure and to support local descriptions and interface terminologies. For these reasons, CKG is further specialized into two areas: archetype governance and terminology governance.

Archetype governance includes the management of existing archetypes and the identification of requirements for creating new archetypes. S. Collins describes a 10 step process for the governance of structured data elements within an EHR implementation [47]: identify clinical topics, create draft models, score data needs, prioritize clinical models based on score, validate models, calculate gap between the EHR and the models, communicate models across project members, request revisions, evaluate usage of the models, and monitor for new evidence requiring revisions. S. Garde describes a similar approach for CKG, but in only 5 steps [45]: archetype development, archetype review, feedback consolidation, publication of archetypes, and post-publication activities.

Those publications consider that archetype modeling is a part of archetype governance, including tasks such as the identification, creation, validation, and monitoring of structured data models. They also include the management of the archetype life cycle and the publication of archetypes as part of the governance process. However, both the archetype modeling and the archetype governance are different processes, although closely related. This is a principle followed in this thesis.

Regarding terminology governance, also known as terminology management, it is the process of maintenance of the contents of terminologies, including the addition or removal of concepts, terms, and relationships. Terminology governance is especially necessary when working with complex terminologies such as SNOMED CT but exceeds the scope of this thesis.

3.2.1. Governance initiatives and systems

Archetype governance heavily relies on tools and specialized platforms to achieve its objectives. In this section, we will analyze existing initiatives and governance information systems to compare their functionalities. Since the number of archetype governance systems available is limited, we will also include other CIM governance information systems in the analysis. The only difference is that they use a different format for defining the clinical information models.

A governance system should provide two main functions:

- To provide **support for storage and publication** of clinical information models. Editing tools should be connected and work in collaboration with the repository, which will support archetype discovery and reuse during the edition process and store the final specification of the model.
- To **facilitate the management** of the governance activities. This includes the control of the life cycle of the models, their versioning, the management of editorial teams, and the management of other documentation and resources related to the models.

In the archetype sphere, most of the existing systems work with openEHR archetypes. The main platform is the openEHR Clinical Knowledge Manager software

(CKM) [48]. It is available as a web page where anybody can explore existing openEHR archetypes, while the archetype editorial team can control their development process (Figure 6). It is probably the most complete information model repository in the world. Additionally, several replicas of the same system serve national or local EHR projects, such as the Norwegian national CKM [49] and the German HiGHmed project CKM [50].

The screenshot shows the openEHR CKM dashboard. The top navigation bar includes 'Archetypes', 'Templates', 'Termsets', 'Release Sets', 'Projects', 'Reports', and 'Help'. The left sidebar lists 'All Resources' and 'Archetypes' with a tree view of categories like Cluster, Composition, Element, Entry, Action, Evaluation, Observation, Instruction, Admin, Section, and Structure. The main content area is divided into several sections: 'Become a Part of Our Online Community' with a 'Register Today!' button; 'What Do You Need to Know?' with links to 'Archetypes', 'Templates', 'Termsets', 'Release sets', 'Projects', and 'Incubators'; 'Quick Search' with a search bar and 'Find Resources' button; and 'Look What We've Been Up To' with a pie chart titled 'Archetypes: Status' showing counts for various stages: Initial / Predraft (199), Draft, Team review, Review suspended, Published (230), Reassess (Draft), Reassess (Team review), and Reassess (Review suspended). On the right, there is a 'News' table with columns for Title, Description, and Date.

Title	Description	Date
Specimen container <i>Updated archetype</i>	Project: Laboratory and pathology testing	10-Sep-2019
Respirations <i>Republished archetype (Patch)</i>	Project: Vital signs	10-Sep-2019
Social determinants of health <i>New project</i>	Collection of archetypes related to measuring or assessing social determinants of health.	10-Sep-2019
Medical Device <i>Republished archetype (Patch)</i>	Project: Common resources	04-Sep-2019
Reference sequence <i>Republished archetype (Patch)</i>	Project: Genomics	03-Sep-2019
Gender <i>Republished archetype (Patch)</i>	Project: Common resources	23-Aug-2019
Symptom/Sign <i>Updated archetype</i>	Project: Common resources	23-Aug-2019
Antenatal - first visit <i>Updated template</i>	Project: MOHW Jamaica	22-Aug-2019
NCD - first visit <i>Updated template</i>	Project: MOHW Jamaica	22-Aug-2019
Triage Assessment <i>Updated template</i>	Project: MOHW Jamaica	22-Aug-2019

Figure 6. openEHR international CKM

As of June 2020, the international openEHR CKM includes around 950 openEHR archetypes, and nearly 125 templates. Figure 7 and Figure 8 show a breakdown of those archetypes and templates. Most of the archetypes are still in a draft status ($n=379$, 40%), while only 109 archetypes (11%) are in a published status. In the case of templates, 81 of 124 templates (65%) are draft, and there are not published 3 templates. However, we must remember that templates are designed for local use, so it is not critical that they are validated and published at an international level.

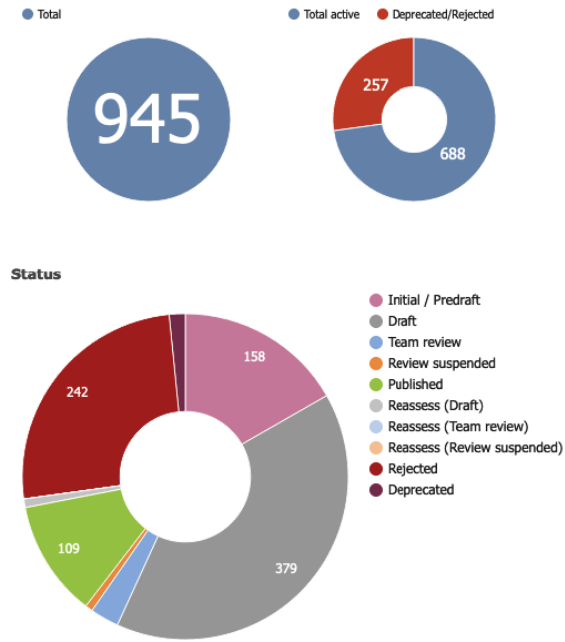


Figure 7. Number of archetypes in the openEHR CKM as of June 2020

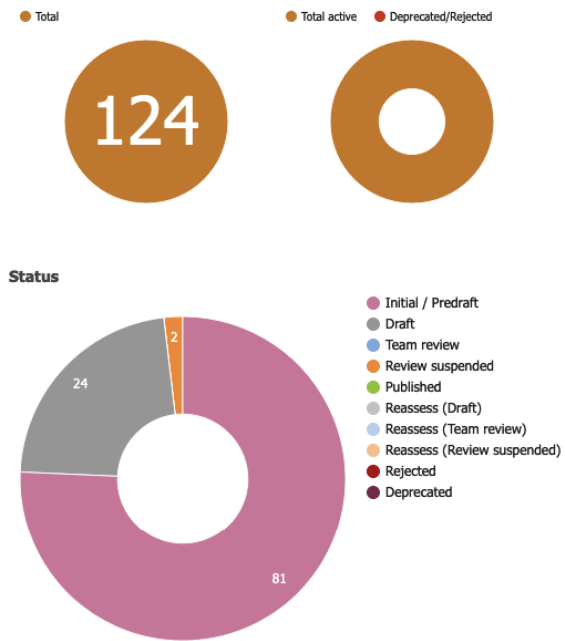
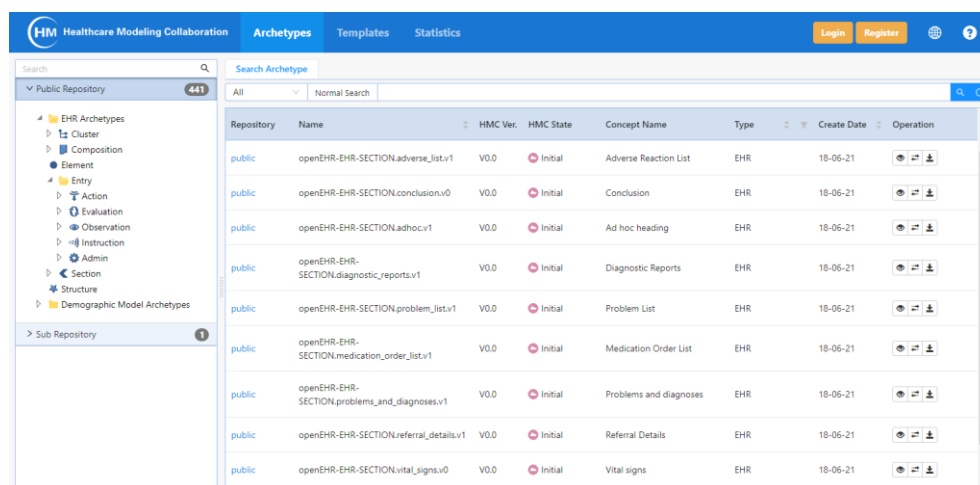


Figure 8. Number of templates in the openEHR CKM as of June 2020

Despite the small number of published archetypes and templates, the achievements of the openEHR CKM is noteworthy. Heather Leslie, one of the main editors of openEHR archetypes, has estimated the time spent in designing, reviewing, publishing and translating all these archetypes in a range between 8.5 and 10 person years [51]. It is remarkable that this is a public, open source effort. It becomes clear that such an effort requires to be done under well established management rules and must have tools to coordinate the participation of multiple people around the world.

China is also in the process of adopting openEHR archetypes. They have developed their own solution for the management of openEHR archetypes, called the Healthcare Modeling Collaboration (HMC) [52]. It is a CKM-like system that supports both openEHR archetypes and templates (Figure 9). As of June 2020, 441 archetypes are available in the platform, plus 22 archetypes and 1 template specific for COVID-19 data.



Repository	Name	HMC Ver.	HMC State	Concept Name	Type	Create Date	Operation
public	openEHR-EHR-SECTION.adverse_list.v1	V0.0	Initial	Adverse Reaction List	EHR	18-06-21	
public	openEHR-EHR-SECTION.conclusion.v0	V0.0	Initial	Conclusion	EHR	18-06-21	
public	openEHR-EHR-SECTION.adhoc.v1	V0.0	Initial	Ad hoc heading	EHR	18-06-21	
public	openEHR-EHR-SECTION.diagnostic_reports.v1	V0.0	Initial	Diagnostic Reports	EHR	18-06-21	
public	openEHR-EHR-SECTION.problem_list.v1	V0.0	Initial	Problem List	EHR	18-06-21	
public	openEHR-EHR-SECTION.medication_order_list.v1	V0.0	Initial	Medication Order List	EHR	18-06-21	
public	openEHR-EHR-SECTION.problems_and_diagnoses.v1	V0.0	Initial	Problems and diagnoses	EHR	18-06-21	
public	openEHR-EHR-SECTION.referral_details.v1	V0.0	Initial	Referral Details	EHR	18-06-21	
public	openEHR-EHR-SECTION.vital_signs.v0	V0.0	Initial	Vital signs	EHR	18-06-21	

Figure 9. Healthcare Modeling Collaboration archetype manager

The Archetype Management System (ArchMS) [53], developed by the Universidad de Murcia, is a system designed for importing, exporting, validating, annotating and searching archetypes using OWL technologies. It can work with both ISO 13606 and openEHR, by mapping them to a common ontology. When an archetype

is uploaded to the platform, it is converted to OWL in order to apply a series of semantic methods. They include archetype constraints validation, archetype transformation between both standards, archetype semantic search, support for archetype-based applications, calculation of archetype similarity, and recommendation of learning contents (Figure 10).

ArchMS

The screenshot shows the ArchMS web application. On the left is a sidebar with three main sections: 'ARCHETYPES' (with a document icon), 'LEARNING RESOURCES' (with a book icon), and 'PROFILES' (with a person icon). The main content area is titled 'Archetypes' and has three tabs: 'Textual Search', 'Advanced Search', and 'Semantic Search'. Below the tabs is a search bar with the text 'Search for:' and a 'Search' button. A table lists several archetypes with the following columns: Name, Archetype ID, Standard, and Life Cycle.

Name	Archetype ID	Standard	Life Cycle
colorectal_screening	openEHR-EHR-EVALUATION.colorectal_screening.v1	OpenEHR	DRAFT
Histopathology	openEHR-EHR-OBSERVATION.lab_test-histopathology.v1	OpenEHR	DRAFT
Histopathology - Specialization: colorectal_screening	openEHR-EHR-OBSERVATION.lab_test-histopathology-colorectal_screening.v1	OpenEHR	DRAFT
Laboratory test	openEHR-EHR-OBSERVATION.lab_test.v1	OpenEHR	DRAFT
Macroscopic findings - Colorectal cancer	openEHR-EHR-CLUSTER.macroscopy_colorectal_carcinoma.v1	OpenEHR	DRAFT

Below the table, there is a pagination control showing '1-5 of 13' and a set of navigation arrows. At the bottom, there is a row of buttons: 'Import an archetype', 'View All Archetypes', 'View ADL', 'Annotate', 'Download ADL', 'Download OWL', 'Download XML', 'Check Archetype', 'Transform Archetype', 'Search Similarity', and 'Search Learning Resource'.

Figure 10. ArchMS system

A special case of study is the openEHR Archetype Designer [54]. This archetype editor can use different cloud data repositories such as Google Drive, Dropbox, or GitHub to store and retrieve the edited archetypes. This fulfills the archetype storage requirement, but does not provide any functionality related to the other governance activities. Thus, this system is not included in our study. A similar case happens with some archetype-based EHR systems. These systems are designed to support data entry and storage of clinical data compliant with archetypes. They usually incorporate an internal archetype repository, but just to support the clinical data management and not the CIM governance itself. Therefore, they are not included in our analysis either. Examples of these systems are ZK-ARCHE [55] and OntoCR [56].

Outside the archetype ecosystem, other CIM repositories exist, which use different formats for clinical information model definitions. We will focus on systems

that provide support for the most common health information interoperability standards based on clinical information models. Those are HL7 CDA, HL7 FHIR, CIMI models and CEM models, all described in Section 2.4.

The most common approach for defining model for the HL7 CDA standard is using implementation guidelines, written documents that describe, in a tabular or semi-structured format, the constraints that a CDA document must fulfill. An implementation guideline is the mechanism for defining HL7 CDA templates.

However, there is a more structured approach for defining and govern these guidelines. It is the ART-DECOR system [57], an open-source tool that supports the creation and maintenance of HL7 CDA templates, value sets, scenarios and data sets (Figure 11). It is a comprehensive system that supports nearly every aspect of the governance. With a solid focus on publishing HL7 CDA template specifications, it does not neglect the management of those specifications and the management of several related resources, such as terminology value sets or the management of projects composed by multiple information models.

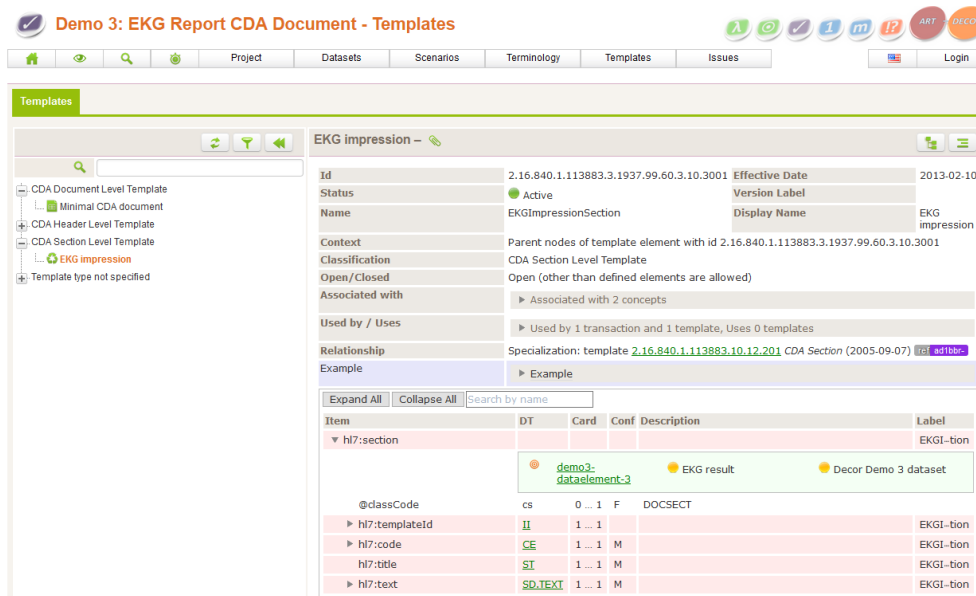


Figure 11. ART-DECOR management system

For HL7 FHIR resources, the main reference is SIMPLIFIER.NET [58] a web site for searching published resources, including profiles, extensions, or terminology value sets. While the HL7 FHIR specifications web page is limited to the publication of the official FHIR resource definitions, SIMPLIFIER.NET is a commercial tool that supports the management of FHIR implementation projects. It is possible to setup a project, select or upload resources to it of any type, assign members to the project with admin, read or write rights, and finally document the project and publish it using a public URL. An example of a FHIR project for cancer information can be seen in Figure 12.

The screenshot shows the SIMPLIFIER.NET website interface. At the top, there is a search bar and navigation links for 'SNIPPET', 'FEEDBACK', and a user profile icon. Below the header, the project name 'NCCN Chemotherapy Order Templates' is displayed, along with a description: 'Templates to improve the safe use of drugs and biologics in cancer care.' The page includes a navigation menu with 'Introduction', 'Resources', 'Guides', 'Members', 'Log', 'Dependencies', and 'Packages'. A 'Resources' section lists various resource types with counts: Profiles (2), ValueSets (7), CodeSystems (5), Extensions (25), and Images (1). The 'Examples' section is currently empty. The 'Canonical Base URLs' section shows 'http://ots.nccn.org/fhir/HL7' with a 'Claimed' status. On the right side, there is a large logo for the 'National Comprehensive Cancer Network' (NCCN) and a paragraph of text describing the project's purpose and its relationship to NCCN guidelines and compendium.

Figure 12. Example of a SIMPLIFIER.NET project

CIMI and CEM are two international initiatives for the creation of generic and interoperable clinical information models. CIMI defines its own RM and uses archetypes for the definition of models, which are published in the CIMI Browser [59]. CEM defines its clinical models using a proprietary format, and they are published in the the OpenCEM Browser [60]. In both cases, the browsers only provide basic search capabilities and a visualization of the models, and cannot be considered complete governance systems.

Finally, we will describe the SOM-SNS governance system. SOM-SNS [61] is the platform used by the Spanish Ministry of Health for the management of CIMs of the National EHR project. The main characteristic of this system is that allows the governance of CIMs independently of any particular RM or CIM format. Using SOM-SNS it is possible to manage ISO 13606 archetypes, openEHR archetypes, HL7 CDA templates, HL7 FHIR resources, and any other definition format, including other documentation needed to explain the underlying knowledge represented in the CIM. It can also manage terminology reference sets. Additionally, both CIMs and terminology reference sets can be managed under the umbrella of a project. Figure 13 shows the representation of a Patient Summary model in SOM-SNS and its related CIM implementations: an ISO 13606 archetype, an ISO 13606 validation script, an HL7 CDA template, a sample data instance, and two SNOMED CT reference sets.

The screenshot displays the SOM-SNS governance platform interface. The top header includes the Spanish Government logo and the SOM-SNS logo. The main content area is titled 'Historia Clínica Resumida' and shows a list of versions. The 'Resources' tab is active, displaying a table of related CIM implementations.

Name	Category	Format	Actions
CEN-EN13606-COMPOSITION.HistoriaClínicaResumida.v1	CEN 13606	adl	[Edit] [Share] [Download] [Delete]
CEN-EN13606-COMPOSITION.HistoriaClínicaResumida.v1	CEN 13606	sch	[Edit] [Share] [Download] [Delete]
HL7-CDA-ClinicalDocument.Documento_HCR.v1	HL7 CDA	adl	[Edit] [Share] [Download] [Delete]
Instancia de ejemplo HCR	CEN 13606	xml	[Edit] [Share] [Download] [Delete]
der2_cRefset_VariablesHCRSpainExtensionExtendedSnapshot	REFSET SNOMED CT	zip	[Edit] [Share] [Download] [Delete]
doc_Refset_VariablesHCRSpainExtensionsSimpleSnapshot_E	DOCUMENTO	pdf	[Edit] [Share] [Download] [Delete]

Figure 13. SOM-SNS governance platform

SOM-SNS is a deployment of LinkEHR Model Manager, a CIM governance system described in detail in Section 6.1.2, as it is part of the results of this thesis.

3.2.2. Comparative analysis

Considering the repositories and systems described in the previous section, we can compare them and analyze their characteristics, functionalities, strengths and limitations. This comparison is made from a functional perspective, since the technologies used to implement the systems does not directly influence in their governance capabilities.

We will limit the analysis to those systems that are actually available and in use by national or international organizations, excluding prototype systems. We will also exclude those systems that only provide a search interface but no additional governance functionalities.

The following characteristics of each of the systems will be analyzed:

- **Supported standards.** Which are the reference models, standard or not, supported by the system?
- **Supported types of CIM.** Which are the formats or types of CIM definitions supported by the system? Archetypes, templates, resources...
- **Accessibility.** Is the system open to public participation?
- **Management of user groups or editorial teams.** Does the system allow the management of user group, assigned to different models or with different responsibilities?
- **Governance of projects or scopes.** Does the system allow to create projects or collections of models focused in a particular scope?
- **Semantic governance.** Does the system use terminologies or ontologies to characterize the models and to support their exploitation?
- **Management of CIM metadata.** Does the system allow adding metadata to the models such as description of the model, authors, participants or purpose?
- **Management of CIM versions.** Does the system control the versioning of the models?

- **Management of CIM lifecycle.** Does the system control the lifecycle of the models?
- **Comments and reviews.** Does the system provide support for commenting or reviewing the models?
- **Translations.** Does the system provide support for translations to multiple languages?
- **Terminological resources.** Does the system support the publication or governance of terminological resources?
- **Technical resources.** Does the system support the publication or governance of technical and implementation-oriented resources?
- **Clinical evidence documentation.** Does the system support the publication or management of domain knowledge literature that support the model's definition?
- **Additional functionalities.** Identification of any additional functionality that is relevant from the governance perspective.

The results of this analysis can be seen in Table 3.

	CKM	HMC	ArchMS	ART-DECOR	SIMPLIFIER	SOM-SNS
Supported standards	openEHR	openEHR	openEHR, ISO 13606	HL7 CDA	HL7 FHIR	Multiple
Supported types of CIM	Archetypes, templates	Archetypes, templates	Archetypes	HL7 CDA templates, HL7 FHIR	All types of HL7 FHIR Re- sources	Multiple
Accessibility	Public participation	Limited to Chinese citi- zens	Public participation	Public or private participation	Public or private participation	Public participation
Management of user groups or editorial teams	Yes	Yes	No	Yes	Yes	Yes
Governance of projects or scopes	Yes	Yes	No	Yes	Yes	Yes
Semantic governance	Yes	Unknown	Yes	No	No	Yes
Management of CIM metadata	Yes	Yes	No	Yes	Yes	Yes

Management of CIM versions	Yes	Yes	Yes	Yes	Yes	Yes
Management of CIM lifecycle	Yes	Yes	Yes	Yes	Yes	Yes
Comments and reviews	Yes	No	No	Yes	Yes	Yes
Translations	Yes	No	No	No	No	Yes
Terminological resources	Yes	No	No	Yes	Yes	Yes
Technical resources	Yes	No	No	Yes	Yes	Yes
Clinical evidence documentation	Yes	No	Yes	No	Yes	Yes
Additional functionalities	Archetype adoption, archetype comparison, distributed CKM	Comparison of archetypes	Similarity search, archetype transformation, archetype validation	Code generation	Additional documentation is provided in a wiki style	Semantic relationships

Table 3. Comparative analysis of CIM governance systems

One of the major conclusions of the analysis is that most of CIM governance systems are only designed to work with one type of CIM format (either archetypes or HL7 artifacts) and using only one RM, or closely related RM at most (openEHR and ISO 13606, or HL7 CDA and HL7 FHIR). There is only one case, SOM-SNS, that covers multiple standards and CIM formats. Being limited to one type of RM and CIM has the benefit of being able to build systems that are specifically designed for them, and that can provide fine-tuned interfaces and functionalities. The drawback is that this approach makes difficult achieving interoperable models between different standards, or reusing the knowledge represented in existing models. Moreover, working with a single RM may fulfill the needs of national or regional EHR interoperability projects, where the decision of the RM to be used is taken for all the territory. In contrast, international EHR modeling initiatives would benefit of being able to accept CIM expressed in any format of following any RM to boost cooperation between their members.

The type of EHR project also influences on the accessibility of the systems. Most of the studied systems are open for public participation, always after a registration process. In some cases, like the Chinese HMC is limited for their citizens. In other cases, the governance systems support both public and private projects, which might be useful for very specific scenarios, or confidential research projects.

A public registration does not mean that the new users have rights to modify anything of the models. Usually, the registration is needed just for audit reasons and provides a read-only access to the models, or the capability of making comments or rising issues about the models. The most common approach in the studied systems is to have a user or role management that allows assigning additional rights to some users. For example, editorial or administrative rights to create new models, to version them, to edit and modify their contents, to change the lifecycle state or even to make them obsolete.

Another functionality that is shared is the possibility of creating projects or domains of use for the CIMs and their related resources. A project can cover a scenario of use, a knowledge area, or a specific work package. CIMS and any other

interoperability resources are then assigned or attached to the project. Working at a project level facilitates the governance of sets of models and any additional terminological, technical or documentation resources. It serves as a single access point to retrieve the interoperability specifications for a particular use case.

We can also take profit of ontologies to offer a semantic governance of CIM. The models can be annotated and classified on criteria like their application domains, by mapping them to a classification ontology or terminology. This creates a semantic description layer over the models that can be used for improved search capabilities. This approach is supported by three out of six of the studied systems, each of them with different capabilities. ArchMS is the system with the most profound support of ontologies as the basis for archetype management. It combines the terminological annotations that are present in the archetypes with ontologies available in BioPortal [62] to create a richer semantic description of the models. OpenEHR CKM uses a predefined internal ontology to classify the uploaded archetypes according their domain of use. Finally, SOM-SNS provides a simple terminological annotation of the models to facilitate their search.

In addition to semantic descriptions, governance should support metadata-based strategies for the identification, description, and control of their evolution. Nearly all the studied systems allow defining metadata of the models. The exception is ArchMS, probably due to their focus on the semantic classification as explained before. In all the other cases, metadata includes the possibility of including information such as a textual description of the model, the context of use, keywords, information about the authors and participants, etc.

It is also important to maintain a version control and a lifecycle status of the models. CIMs are entities in constant evolution. They are created in an iterative process, refining their information structure and terminology mappings. Version control allows a seamless coexistence of different definitions of the models. Each version of the model must have its own identifier, and each version can be retrieved from the system in case it is needed. Lifecycle status serves to know the reliability

of the models (from a draft model, to a published or obsolete model). All the studied systems support this kind of information.

Collaboration in the creation and evolution of CIMs is part of the essence of governance systems. For that reason, most of the studied systems include the possibility of adding comments to the models or raising concerns about them. A powerful community of users is essential to define quality CIMs since they will provide different points of view and requirements to be fulfilled by the models. Governance systems must provide mechanisms to promote this participation, as it occurs in most of the studied systems. In this aspect, openEHR CKM is probably a referent for all the other systems given its vivid community of archetype editors and users.

The support of translations to multiple languages is one of the negative results of this comparative study. There is a clear tendency to limit the description and metadata of the CIMs to a single language, usually English. Although governance systems will allow using a different language for the metadata, they do not usually allow the coexistence of multiple languages. Here we are not talking about the actual contents of the CIM, and its information structure, since the multilingual support will depend on the characteristics of each CIM format, but to the multilingual support of the metadata accessible using the governance system interface. There are two significative exceptions to this, the openEHR CKM and SOM-SNS. Both systems implement a native support for multilingual metadata, which mean that all the descriptions and documentation of the models can be translated to multiple languages, empowering the international collaboration in the edition and reuse of the information models.

We finally have to analyze the support of additional artifacts or resources attached to a CIM definition. Table 3 distinguishes three types of resources: terminological resources, technical resources, and clinical evidence resources.

Terminological resources refer to value sets, subsets, reference sets, or any other specification format of collections of terminology codes. As we know, a CIM is a combination of an information structure and a semantic description of that in-

formation by means of terminologies. When a CIM is defined, it should be accompanied by terminology sets that indicate the acceptable coded values in some specific data elements. Arguably, these terminology value sets should be managed by a dedicated system: a terminology server. However, it is also useful that the CIM governance system provides a basic management of such value sets or, at least, the ability of linking the CIM to a terminology server. OpenEHR CKM, ART-DÉCOR, SIMPLIFIER.NET and SOM-SNS include functionality for the terminological resources management. They can store named collections of codes or reference to existing ones, and in some cases, such as SOM-SNS, it is possible to create intensional reference sets, i.e. dynamically create the value set by defining the properties that must be fulfilled by the codes.

Technical resources are all the artifacts that can accompany a CIM and that support its implementation in EHR systems. Technical resources can be of a wide variety: sample data, programming scripts, implementation guidelines, prototype screens generators, etc. The objective of these resources is to help minimizing the gap between the medical definition of a CIM and knowledge of technical implementers. OpenEHR CKM, SIMPLIFIER.NET and SOM-SNS support technical resources management. ART-DÉCOR does not allow attaching technical resources, but it is able to automatically generate code to support implementation activities.

Clinical evidence or domain knowledge resources are the documentation that validates the data contents of the CIM structure. It is mainly composed by medical literature and examples of how existing EHR systems are representing a particular information. The management of clinical evidence documentation is made differently in each of the systems that support it. While openEHR CKM and SOM-SNS allow attaching any type of document or medical literature related to the model design, SIMPLIFIER.NET provides an embedded wiki system to document any desired aspect of the model. The most advanced proposal comes from ArchMS, which can recommend learning contents related to an archetype that may help patients or clinicians to improve their knowledge about a specific health issue.

Chapter 4

Archetype modeling methodology

This chapter is an excerpt of the paper “Archetype modeling methodology” by D. Moner, J.A. Maldonado and M. Robles, originally published in the Journal of Biomedical Informatics, 79 (2018) 71–81. [63]

The methodologies described in Section 3.1 were used as an initial reference for our development. Learning from existing good practices in modeling clinical information models, information structures, and ontologies helped in setting the initial main phases of our archetype modeling methodology. This helped to align our archetype modeling methodology to the modeling of other health information artefacts. Since all these artefacts work coordinately inside information systems, it is recommendable that they follow the same design principles.

4.1. Requirements

We define the methodology based on the specific needs and characteristics of archetypes [64], and their development process. We present next a comprehensive list of requirements covered by the methodology.

- Archetypes are built according to an underlying RM. Archetype authors must consider the underlying structure and the contextual information already supported by the RM when creating archetypes.
- Ensure that the archetypes represent all or most of the information required in the proposed scenarios of use.
- Facilitate the creation of reusable archetypes. An archetype definition should be sufficiently generic to be reusable in other scenarios. Afterwards,

archetypes can be modified by specialization and versioning in order to meet specific requirements.

- Coordinate the concurrent use of archetypes and terminologies. Terminologies are used to define the semantics of the archetype structure itself, and to define the valid values for coded data. Guidance on how to combine archetypes and terminologies should be provided.
- Support the participation of technical and clinical specialist in the modeling process. Building archetypes is a multidisciplinary task. It requires the participation of experts on the medical domain and experts on the technical standards used. An archetype development methodology should define the role of the participants.
- Guarantee archetype quality [65]. *Menárguez-Tortosa et al.* [66] analyzed the quality of existing archetypes and concluded that “around 1/5 of archetype specializations contain modeling errors, the most common mistakes being related to coded terms and terminological bindings. [...] This result reinforces the need for making serious efforts in improving archetype design processes”.
- Establish the relationship between archetype modeling and archetype governance. Archetype governance includes the set of policies, actions, and tools to ensure that archetypes can be identified, stored, searched, and that they can evolve to meet new requirements. The relationship between archetypes, and between archetypes and other semantic EHR resources (terminology value sets or clinical guidelines for example) has to be properly managed to guarantee an ecosystem of quality interoperability assets [67,68]. Although this paper does not cover archetype governance, it wraps our archetype development methodology proposal.
- Templates are particular types of archetypes for specific use cases. While archetypes provide generic and reusable definitions of the information model, templates define specific configurations of archetypes. The definition of templates must be covered by the methodology.

The methodology described in this section is the result of nearly 15 years of practical development of archetypes in regional and national EHR projects. The development of the methodology is based on an iterative trial and error approach. Each implementation experience served to learn about the limitations of the methodology that triggered the definition of improved versions.

The **Archetype Modeling Methodology (AMM)** formally defines the common phases and good practices to follow when developing archetypes. The methodology also covers two fundamental aspects for the success of archetype development: the selection of the group of people participating in the modeling process, and the need of using design guidelines to guarantee the consistence of the outcomes.

4.2. Work group

Success of archetype modeling depends on the group of people involved in the process. Archetypes can be used in multiple settings, potentially serving multiple healthcare specialties. A good archetype design team should include, at least, health professionals providing knowledge about different domains of use, experts in clinical terminologies, and technical professionals who are familiar to the RM and the archetype tooling. The following members should be part of the archetype modeling group:

- **Group leader.** Person in charge of coordinating the work of the group and responsible of governing the archetype modeling process. Preferably, the group leader should have a clinical profile, with also some technical skills and knowledge of the archetype development process. Archetype modeling requires reaching a consensus and harmonization of different needs and opinions. The group leader has to make a decision when discrepancies arise between the team members, to facilitate agreements between them.

- **Clinical experts.** They responsible for providing inputs to the modeling process. They have to define the scope of the archetype, to collect information requirements, to document the sources of knowledge and references used, and to select the information items to be included in the archetype. In addition, they decide the structural organization of the information items, and the applicable data constraints.
- **Terminology experts.** Optionally, terminology experts provide inputs related to the use of terminologies in combination with archetypes. They are in charge of defining the semantic binding between the archetype structure and terminology codes. In addition, terminology experts are in charge of defining value sets used in coded information elements.
- **Technical experts.** Optionally, technical experts with expertise in EHR standards and EHR systems implementation can provide insights on the existing information systems. They also provide advice about difficulties or possible limitations for the implementation of the archetypes using a specific RM, and they can develop the final formal specification of archetypes.
- **Multidisciplinary clinical support team.** Additionally, a multidisciplinary group of clinicians from different knowledge areas should collaborate with the work group. They support the information gathering process and the evaluation process of the resulting archetypes, although they might not directly participate in the modeling process.

4.3. Selection of the Reference Model

In order to work with archetypes, it is essential to choose the RM that will be the basis for their definition. Only ISO 13606 and openEHR work natively with archetypes. However, it has been demonstrated that it is possible to apply the dual model methodology to other standards, such as HL7 CDA or CDISC ODM [69]. Each of the RMs has its own technical characteristics and scope. The needs of the working scenario have to be studied to select the most appropriate standard in each case.

However, it is possible that the RM is already decided by the local or organizational regulations.

4.4. Design guidelines

A design guideline describes the general rules, best practices, and common agreements to ensure a consistent development of archetypes at a national, regional or organizational level. Archetype design guidelines specify common rules on how to name archetype nodes, how to create commonly used data structures, or how to specify the preferred reference terminologies. Guidelines may also describe the policy followed to achieve consensus during the development of archetypes. An example of an archetype design guideline is provided by openEHR [70]. This guideline describes what information is expected to be provided for the archetype metadata and the editorial style to be followed.

The use of appropriate tools such as archetype editors [69,71] can support the development process of archetypes and facilitate the achievement of a consistent modeling. Following design guidelines facilitates the maintenance, reuse and interoperability of archetypes in broader contexts. The group leader is in charge of developing new guidelines or selecting existing ones. Members of the work group should be familiar with the guidelines before defining archetypes.

4.5. Methodology phases

AMM is a 5-phase methodology that covers the modeling process of archetypes, from the requirements analysis, to the publication of the result archetypes. The phases are analysis, design, development, validation, and publication. The phases are divided into activities. Each activity includes a description of the tasks, the needed inputs, expected outputs, the tools used, and the participants. The only specialized software needed is an archetype editor, which facilitates the archetype implementation phase.

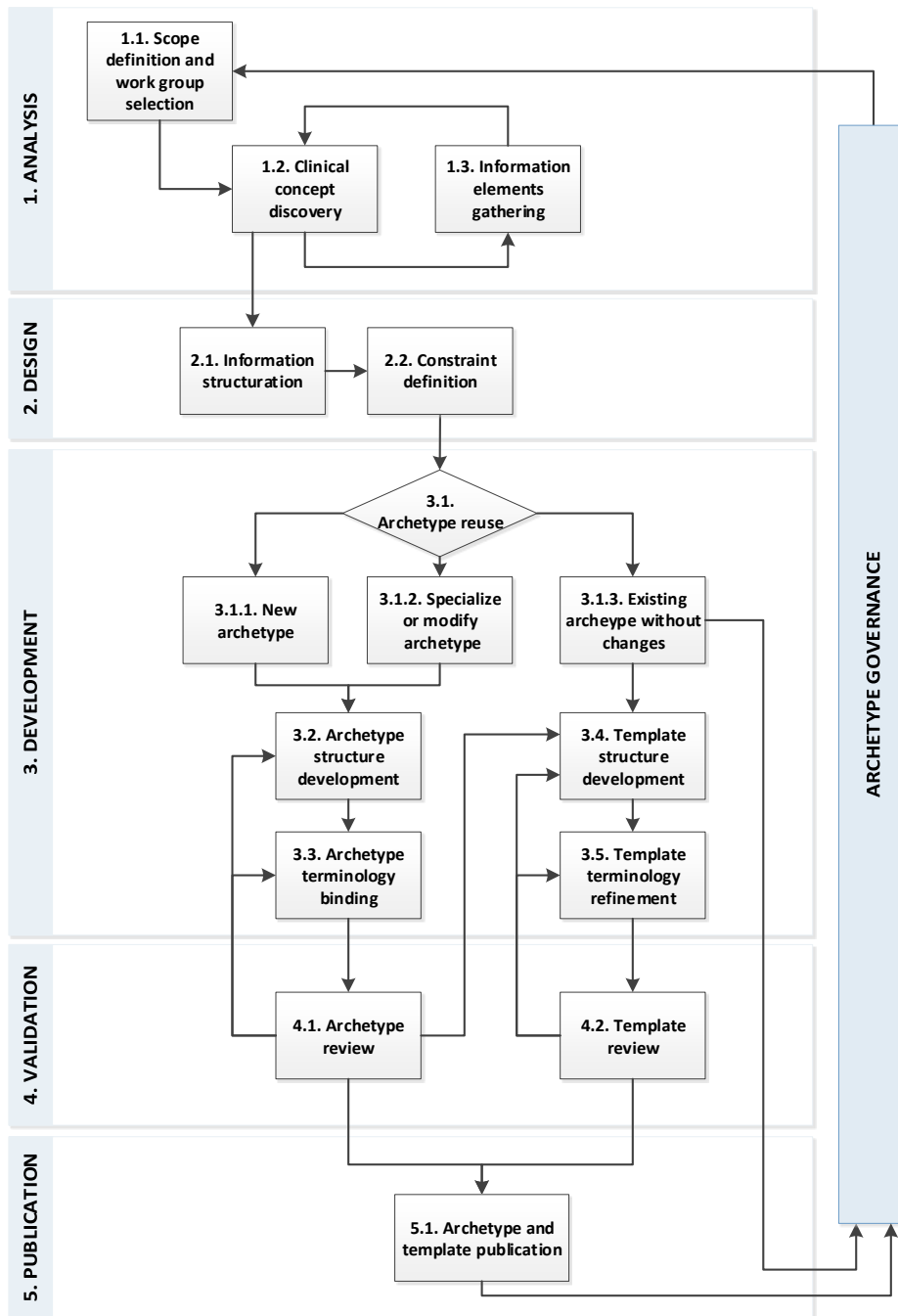


Figure 14. Graphic summary of the Archetype Modeling Methodology workflow

4.5.1. Phase 1. Analysis

The modeling process starts when a promoter, usually a health professional or organization, requires building archetypes to support the information used or registered in a specific domain. The objective of the analysis phase is to delimitate the scope, requirements and use cases of the clinical domain to be modeled. An additional objective is to identify the coarse-grained clinical concepts involved, and gather relevant information for the design phase. Figure 15 shows a simplified example of the results of this analysis phase in the form of a mind map representing medication information.

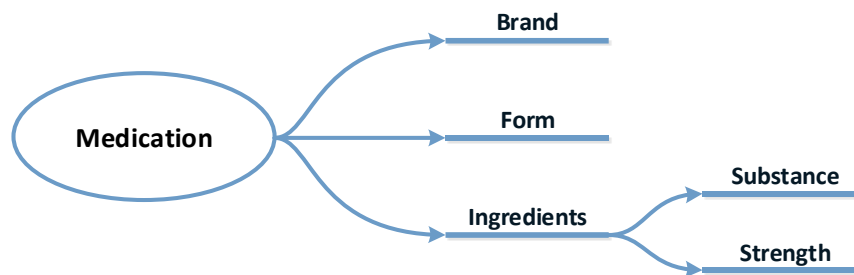


Figure 15. A mind map can easily represent the set of concepts and information elements discovered during the analysis phase

- **Activity 1.1. Scope definition and work group selection**
 - **Inputs:** A request for modeling clinical information in a specific domain.
 - **Outputs:** Initial analysis document. It includes, at least, the scope of the work, the expected uses of the information, the involved information systems and care settings, and the list of members of the work group.
 - **Description of the activity:** The first task is the definition of the precise scope of the archetypes. A limited scope may result in a set of archetypes only usable in a very particular scenario. A broad scope may end in a large set of archetypes defined simultaneously, hindering the overall definition process. The promoter and the leader of the modeling work have to define precisely the limits of the scope and use cases to

be covered. Then, the members of the work group are selected with a multidisciplinary perspective. The participation of experts in the domain of study is essential to gain specialized knowledge about the requirements, but it is also important to incorporate experts from other fields that can offer new perspectives over the problem under study. Finally, it is convenient to provide a first list of information systems and care settings where the archetypes are expected to be used.

- **Participants:** Promoter of the modeling process, group leader.
- **Tools:** Text editor.
- **Activity 1.2. Clinical concept discovery**
 - **Inputs:** Document with the analysis of the scope, requirements, and use cases for the archetypes.
 - **Outputs:** Document with the list of clinical concepts involved in the scenarios of use, including the name of each clinical concept, and a short description of it.
 - **Description of the activity:** The objective is to provide a name and describe all the clinical concepts that are relevant for each use case. Clinical concepts are generic groups of related information involved in the modeled scenarios of use. They do not necessary have a one-to-one match to the archetypes designed in the next phase, which may be influenced by the internal architecture of the selected RM. Multiple archetypes can potentially be derived from a single concept during the design phase. The set of identified clinical concepts must cover the complete scope and requirements defined in activity 1.1.
 - **Participants:** Group leader, clinical experts.
 - **Tools:** Text editor, spreadsheet, mind map.
- **Activity 1.3. Information elements gathering**
 - **Inputs:** Document with the list of clinical concepts.
 - **Outputs:** Document with the list of information elements associated to each clinical concept.

- **Description of the activity:** The objective of this activity is to collect the list of specific information elements associated to each clinical concept. Information elements are atomic clinical data items registered in the EHR or used in the modeled scenario. Clinical experts' knowledge and expertise is essential to analyze existing documentation, bibliography, data entry forms, data interchange messages, data structures (databases, information systems, etc.) related to the scope of each clinical concept. They have to decide which data is needed at different healthcare levels (primary care, specialized care, emergency care, clinical research, etc.), or requested by different users or roles (clinicians, nurses, researchers, administrative staff, etc.). The result will be a collection of information elements that will be part of the archetype definition.
- **Participants:** Group leader, clinical experts.
- **Tools:** Bibliography, documentation of existing information systems, text editor, spreadsheet, mind map.

4.5.2. Phase 2. Design

Once the scope of the modeling process is clear, and the clinical concepts and information elements involved are identified, the information has to be organized and structured. The constraints on data values are also defined. Figure 16 shows the simplified design table for the medication archetype.

- **Activity 2.1. Information structuration**
 - **Inputs:** Document with list of clinical concepts and information elements.
 - **Outputs:** Document with the archetype structural design.
 - **Description of the activity:** An aggregation of the information elements is needed to create meaningful archetypes. Aggregations might correspond, or not, to the originally identified clinical concepts. For example, two different clinical concepts may not have much sense as

standalone archetypes and may be modeled as part of the same archetype. Each designed archetype will have a name, purpose, the list of information elements included, and how they are structured and organized. It is recommended, although not mandatory, to be as much aligned as possible to the RM chosen for the final implementation. Each RM imposes a basic structure for the data with a particular meaning and purpose. For example, ISO 13606 defines Compositions, Sections, Entries, Clusters and Elements as the basic building blocks of the information structures. Data structures should be compliant with the underlying RM. It is also possible to design the archetype in a generic way, without specifying RM classes, and only include the RM structures during the following development phase. The results of the activity will be the main reference to build the final archetypes.

- **Participants:** Group leader, clinical experts.
- **Tools:** Text editor, spreadsheet, mind map.

Archetype description					
Name	Medication				
Description	Information about a medication				
Recommended use	Used together with medication prescription and administration instructions				
Leader					
Participants					

Archetype design					
Information element	Description	Mandatory	Repeteable	Class/Data type	Domain
Medication name	Commercial brand	Yes	No	Free or coded text	
Generic name		Yes	No	Coded value	SNOMED CT
Form	Presentation. E.g. tablets	Yes	No	Coded value	SNOMED CT
Active ingredients	Add one group per each ingredient	No	1..*	Organizer	
Substance name		No	No	Free or coded text	
Strength		No	No	Physical quantity	>0, UCUM
Additional details		No	No	Free text	

Figure 16. Example of a design table created during Phase 2

- **Activity 2.2. Constraint definition**
 - **Inputs:** Document with the archetype structural design.
 - **Outputs:** Document with the complete archetype design specifications, including data constraints for each information element.

- **Description of the activity:** The archetype structure is refined by including information about the constraints applicable to data elements, such as occurrences, cardinality, or data types of each element. Data constraints applicable to data values are also specified: ranges for numerical values, valid alternatives or fixed values for texts and string values, or lists of terminologies that can be used to record data for coded values. The constraints and classes defined by the underlying RM could be considered, for example to decide the data types of each information element.
- **Participants:** Group leader, clinical experts.
- **Tools:** Text editor, spreadsheet, mind map.

4.5.3. Phase 3. Development

The development phase, is focused on the creation of the actual archetypes, using the appropriate technology and tools, such as the Archetype Definition Language (ADL) [72], archetype editors, and terminology services. New archetypes might be created, or existing ones might be reused or adapted if needed. Local configurations of archetypes (templates) might be also created in order to fit specific requirements. Figure 17 shows a medication archetype, as the result of the development phase.

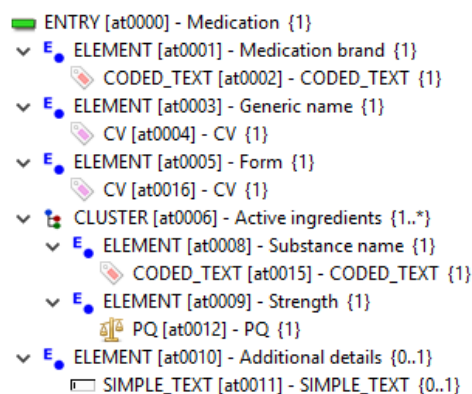


Figure 17. Information structure of a medication archetype implemented in ISO 13606

- **Activity 3.1. Archetype reuse**

- **Inputs:** Document with the complete archetype design specifications.
- **Outputs:** List of existing reusable archetypes, either completely or needing modifications, and new archetypes to be developed from scratch.
- **Description of the activity:** The aim of this activity is to select the set of reusable existing archetypes, and list the new ones to be defined. Archetypes related to the scope of our modeling scenario were probably studied during the phase of Analysis, since they are also a source of information and knowledge. It is probable that some existing archetypes fit in our information models with no additional changes (activity 3.1.3). If it is not the case, new archetypes can be created (activity 3.1.1), or adapted by specialization or versioning (activity 3.1.2). Not all the information present in the archetype design specification may become part of the final archetype, as the RM usually already contains most of the common contextual information related to the health care process. Technical experts will identify which information elements do not have to be represented explicitly in the archetype structure.
- **Participants:** Group leader, technical experts.
- **Tools:** Archetype repositories.

- **Activity 3.2. Archetype structure development**

- **Inputs:** Document with the complete archetype design specifications, list of existing and reusable archetypes, and list of new archetypes to be developed.
- **Outputs:** Set of archetypes implemented in a formal archetype definition language.
- **Description of the activity:** This activity is executed for each of the archetypes to be developed or adapted. Archetypes can be implemented using a specific formal language such as ADL. It facilitates defining all the constraints and structure of an archetype using a well-defined syntax. To ease the implementation process, specific tools such as

an archetype editor can be used. In this activity, it is required to be strictly compliant with the RM selected. Building an archetype starts by choosing its root class. The root class will limit the valid nested information elements, which will be structured and constrained according to the RM. This includes deciding about the best matching between the information model specifications and the classes and structures imposed by the RM. All constraints over data values have to be included in the archetype implementation. In addition, a name and description of all information elements using natural language has to be provided. The metadata of the archetype (authors, expected use, keywords, etc.) should be also defined. Clinical experts trained in the use of archetype editors and knowledgeable about the RM can perform the structure specification activity. However, given the complexity of health standards and RMs, it is recommendable that technical experts in the RM participate in the development.

- **Participants:** Clinical experts, technical experts.
- **Tools:** Archetype editor.
- **Activity 3.3. Archetype terminology binding**
 - **Inputs:** Set of implemented archetype information structures.
 - **Outputs:** Set of archetypes bound to terminologies.
 - **Description of the activity:** Medical terminologies provide the semantics of archetypes. Archetypes are bound to terminologies in two ways. First, the model meaning binding or term binding, where a descriptive code is assigned to each information element of the archetype. Second, the value set binding or constrain binding, where valid value sets can be assigned to coded information elements. It is usually necessary to use specialized systems such as terminology servers, to help in building and managing value sets. An in-depth description of the work to be made with terminologies and how to create usable value sets aligned to archetypes is out of scope of this paper, but there are many references in the literature [73–77].

- **Participants:** Terminology experts, clinical experts, technical experts.
- **Tools:** Archetype editor, terminology services.
- **Activity 3.4. Template structure development**
 - **Inputs:** Set of archetypes.
 - **Outputs:** Set of templates.
 - **Description of the activity:** Templates are a special type of archetypes that further constrain them for a particular setting. In other words, templates are archetypes configured for a local use, and not intended for general reuse. A template represents the most specific requirements of the initial scenarios of use. Templates focus on the usability of the information model rather than in its genericity and interoperability. The process of defining templates includes selecting existing archetypes and put them together in a broader information model (usually at the level of clinical documents). In a template, information elements that are not needed in a local scenario are removed. Other information elements may be further constrained.
 - **Participants:** Clinical experts, technical experts.
 - **Tools:** Archetype editor.
- **Activity 3.5. Template terminology refinement**
 - **Inputs:** Set of templates.
 - **Outputs:** Set of templates with refined terminology bindings.
 - **Description of the activity:** Templates can also refine terminology bindings. For example, in a particular template, a subset of the valid codes for an information element can be defined. Standard terms can be adapted to local vocabularies in use by the final users and systems.
 - **Participants:** Terminology experts, clinical experts, technical experts.
 - **Tools:** Archetype editor, terminology services.

4.5.4. Phase 4. Validation

The aim of the validation phase is to ensure that archetypes and templates meet the initial needs and requirements. The validation phase includes identifying

errors, inconsistencies, absences of information, or misleading specifications. If errors are found, archetypes or templates have to iterate again over the previous development phase.

- **Activity 4.1. Archetype review**

- **Inputs:** Set of archetypes.
- **Outputs:** Set of validated archetypes or list of needed changes. It will require a new iteration of the development activities.
- **Description of the activity:** Automatic review and validation of archetypes is possible to some degree. It includes the validation of the information structure with regard to the underlying RM, and the consistency of terminology bindings [66]. However, an additional functional validation should be performed to check if archetypes are fit for purpose. Final users of the archetypes should be responsible of the functional validation. To facilitate the validation activity, technical complexity of archetypes and RMs models should be hidden as much as possible. This can be achieved by using alternative archetype representations, automatically generated from the specification, such as mind maps or data entry forms. If the participation of a selected group of final users is not possible, the multidisciplinary clinical support team can be responsible of the validation activity. Every comment or error notification about the archetypes should be registered, tracked and adequately resolved before deploying the archetype.
- **Participants:** Group leader, clinical experts, multidisciplinary clinical support team, selected final users.
- **Tools:** Archetype editor, issue-tracking system.

- **Activity 4.2. Template review**

- **Inputs:** Set of templates.
- **Outputs:** Set of validated templates or list of needed changes. It will require a new iteration of the development activities.

- **Description of the activity:** This activity is equivalent to activity 4.1, but related to the review of templates. The same recommendations apply here, although the participation of end users is even more important, since templates will be closer user interfaces, and to the real implementation inside EHR systems.
- **Participants:** Group leader, clinical experts, multidisciplinary clinical support team, selected final users.
- **Tools:** Archetype editor, issue-tracking system.

4.5.5. Phase 5. Publication

The last phase of the methodology is the publication of archetypes and templates. The objective is to make them available for EHR systems developers, and to facilitate their reuse for creating new or specialized archetypes. Publication of archetypes and templates is closely related to archetype governance.

- **Activity 5.1. Archetype and template publication**
 - **Inputs:** Set of archetypes and templates.
 - **Outputs:** Published archetypes and templates.
 - **Description of the activity:** Publication of archetypes and templates means making them available to any user or system. The formal definition of archetypes (ADL) should be downloadable, editable and reusable. In some cases, the developers may apply restrictive or commercial licenses to archetypes. In that case, archetypes cannot be considered as part of an interoperable ecosystem, but only as a documentation of a particular implementation. Archetypes and templates in development (drafts) can be published in order to open them to a public review process, but in that case they have to be properly identified [78] and users should be warned about the potential risks of using those definitions in a real implementation.
 - **Participants:** Group leader.
 - **Tools:** Archetype repository.

4.6. Discussion

In this chapter, we have presented a formal methodology to guide the definition and implementation of archetypes. The adoption of archetypes and dual model architectures is entering into its maturity. However, in order to achieve a robust and stable archetype-based environment, it is fundamental to have a clear modeling methodology.

Archetype-based semantic interoperability requires using standard RMs, and medical terminologies to attach semantic descriptions to information structures. This would be enough to guarantee the faithful exchange and reuse of data and models among heterogeneous systems in a full semantic interoperability scenario. However, the lack of a homogeneous methodology for defining archetypes, and for governing them, burdens real semantic interoperability. Without such a methodology, different people or organizations might create different and incompatible archetypes for the same purpose. Archetypes might overlap or leave out portions of the information, limiting their reusability.

This work provides a standard methodology for archetype development. In order to measure the success of the methodology, we can analyze how well it covers all the aspects related to the development process, how much it follows established good practices, and which degree of adoption it reaches.

AMM can be compared to existing development methodologies of related areas, specifically, methodologies for the development of CIMS, software and ontologies.

1. Analysis. All methodologies include a first phase related to the analysis of the general domain, the study of the specific use case, and the identification of requirements. In the case of development of CIMS, it includes the clear definition of the scope of use of CIMS. It also includes the analysis of the domain to acquire information of the existing systems and solutions. In the case of software development, the requirements are technical (related to

implementation and performance issues) and functional (related to the inputs, outputs and behavior of the system). In the case of ontologies, the requirements phase establishes the objective and scenarios of use of the ontology. Finally, in AMM, this phase also includes the identification of clinical concepts and information elements.

2. Design. The goal is to completely describe and organize the contents identified during the analysis phase. The design phase of AMM, CIMs and ontologies is very similar. In software development, design is slightly different, since it covers three aspects: architecture, functional and technical design. It includes not only the information specifications, but also the expected behavior of the software and the global context of execution. In other words, software design covers a bigger scope than the design of archetypes.
3. Development. It is the process of actual development of the archetypes in a formal language. In the AMM it includes the development of the information structure and constraints using ADL language. The development phase also includes the selection of the controlled vocabularies (i.e. terminologies or value sets) used by the models. CIMs definition phase has similar objectives. There are more differences with software development phase. The result of archetype development is a type of specification expressed in a formal language computable by EHR information systems. In contrast, the result of software development is the working system itself.
4. Validation. The objective of the validation phase is to confirm that the specified model covers all requirements and use cases previously identified. It is a common phase for all the mentioned methodologies, as it can detect errors generated during the development phase, or even requirements not considered initially.
5. Publication. The publication phase has some differences between the studied methodologies. In AMM and CIMs development, it consists in publishing archetypes and templates (or CIMs) to make them available for systems that can use the new models. In contrast, a deployed software or ontology is ready for use by the final users.

In summary, AMM phases align to other methodologies. Table 4 shows a summary of this comparison. There are, of course, differences related to the specific aspects and nature of archetypes, CIMs, software, and ontologies, but they do not affect the overall approach of the methodology.

Archetype modeling methodology	Development of clinical information models	Development of software	Development of ontologies
Analysis	Scope definition Selection of experts Analysis of domains	Identification of needs Functional requirements Technical requirements	Specification Knowledge acquisition
Design	Design	Architecture design Functional design Technical design	Conceptualization
Development	Definition	Implementation	Implementation
Validation	Validation	Testing	Evaluation
Publication	Publication	Installation	Application

Table 4. Comparison of AMM, CIM, software and ontology development methodologies

Experts involved in the development of archetypes influence their final quality. For this reason, we have provided recommendations about the members of the work group. It is important to have a multidisciplinary group of clinical experts to

ensure the genericity of archetypes. They will provide clinical documentation requirements for each of their specialties. It is also important to reach an agreement on the minimum common aspects and information elements included in the archetypes. Work group members should be trained in clinical information modeling and know about the benefits of building generic and reusable models. It is essential to maintain a stable group of trained professionals in archetype modeling, and to involve them in future developments or revisions of the archetypes, to maximize the quality of archetypes.

It is important to define a realistic scope for the modeled archetypes (activity 1.1.). Archetype modeling is a costly task. In many occasions, it may be preferable to define a basic but usable archetype, rather than trying to be exhaustive including information elements that cover any clinical scenario. Archetype versioning allows us to define basic versions of the archetypes first, and then review and complete them in future iterations of the modeling process.

AMM does not impose a specific order to act when there is a need to develop multiple archetypes. In that case, we can follow three possible approaches, namely top-down, bottom-up, and middle-out.

In the top-down approach, the most generic or coarse-grained concepts and information structures of the domain under study are first identified. They are analyzed to find specific structures that can be reuse in different places. The main problem with the top-down approach is that it usually leads to archetypes dependent on the requirements of the domain of study. Archetypes are usually closer to specific local requirements, i.e. they solve problems related to specific scenarios of use, as for example the definition of the contents of clinical documents or data entry screens. Moreover, in the top-down approach designers do not usually pay much attention to the future reuse of the archetypes.

The second approach is bottom-up. It consists on identifying the most basic information structures inside the EHR, and then composing complex structures re-using the basic ones. Bottom-up approach ensures a higher level of interoperability of the archetypes, since it starts designing generic and basic structures that serve

to any purpose. However, it can also cause the creation of a large number of archetypes, which are not needed in other scenarios. Moreover, it is difficult to identify basic information structures without knowing the context where they will be used.

The third possibility is to follow a middle-out approach. It is a combination of the top-down and bottom-up approaches. In this approach, clinical statements needed in our case of study or usable in other scenarios are identified. A clinical statement is as a building block of the EHR that puts together the information that a human or a machine records, related to a common clinical context. I.e. what has been done, by whom, when, where and how it was done. It corresponds to the ENTRY class in ISO 13606; the OBSERVATION, EVALUATION, INSTRUCTION, and ACTION classes in openEHR; and to the clinical statements of HL7 CDA: Act, Observation, SubstanceAdministration, Supply, Procedure, Encounter, and Organizer. Once the clinical statements are identified, we can proceed downwards, defining the details of the information structure; or upwards, defining templates that reuse clinical statements in configurations for specific scenarios of use.

Although all three approaches can lead to successful results, our experience suggests that the middle-out approach is the most efficient. It provides a good balance between genericity and specialization. Modeled archetypes are sufficient to represent all kind of information structures, without losing their reusability capabilities. Moreover, the middle-out approach helps in distinguishing archetypes for generic scenarios of use, from templates designed for local uses.

One of the main challenges of archetype modeling is to manage the relationship to medical terminologies. During the design phase, the boundary problem between the information model (archetypes and the RM) and the knowledge model (clinical terminologies) may appear. The same information can be often represented as part of the archetype structure or as terminology concepts. Both can be correct representations of the clinical information models, and the decision made will affect how the registered information in the EHR systems can be reused. As a recommendation, AMM requires the definition of the archetype information struc-

ture (activity 3.2) before the definition of the terminology specifications and bindings (activity 3.3). Advanced terminologies, such as SNOMED CT, include a complete conceptual model, based on formal description logics, constraining the permitted attributes and values that may be applied to each kind of concept. However, this conceptual model is not enough to represent all the epistemological and contextual information to be documented during health care. EHR data standards and archetypes provide the basic framework to represent clinical data and its context information. They also define where terminologies should be used, and thus, they help defining the scope of terminology value sets. Value sets might not be aligned to data structures if their order of definition is swapped.

Archetype modeling is a process executed in the context of a broader governance process. Archetype governance is responsible, among other tasks, of maintaining published archetypes available, receiving change requests or new requirements, deciding about the deprecation or obsolescence of archetypes, or starting a new archetype modeling process. Chapter 5 describes in detail the archetype governance process.

AMM methodology describes the archetype development process. However, the AMM methodology can also serve as a basis for the development of any other type of CIM, for example HL7 CDA templates or HL7 FHIR resources. Different technologies and tools may be used in each case, but the principles of analysis, design development, validation, and publication will remain the same.

The lack of a proper archetype modeling methodology limits the definition of quality-assured and interoperable archetypes. The proposed AMM methodology has taken into account the best practices from other development domains and adapted them to the archetype characteristics. Our aim is to provide a reference methodology. Archetype designers and developers can adapt the proposed phases to meet some specific needs of a particular development project. For example, personal or technical resources involved in modeling a national EHR would not be the same as for modeling a local information system. However, they should follow the same basic design principles.

Chapter 5

Archetype governance

This chapter provides a definition of archetype governance, its responsibilities and duties, the areas of work, and the involved personnel. It does not describe a strict methodology, since archetype governance might depend on organizational policies and established IT management procedures. However, we can provide a set of requirements and good practices that any health organization should observe to implement an archetype governance process.

This chapter is the result of the analysis of the literature about CIM and archetype governance systems referenced in Section 3.2. Following those analyses, we have defined the scope, basic characteristics and strategies that should be covered by any archetype governance system.

5.1. Definition and scope

As explained in Chapter 4, archetype modeling takes place in the context of archetype governance. Both processes have a different scope but must be coordinated.

We define archetype governance as follows:

Definition 1. *Archetype governance is a framework that comprises the set of processes and policies, within an organization or supra-organization, which encompasses and guides archetype modeling activities.*

Archetype governance is usually linked to a specific semantic interoperability project, defined as an initiative to achieve semantic interoperability among different information system. The extent of this project can vary from the development of a solution for the communication of data in a specific clinical scenario, to

the definition of the archetypes that support the complete semantic interoperability needs of a country or for an international health information communication scenario. Archetype governance is the means to guarantee that the archetypes developed in a semantic interoperability project fulfill the project information requirements.

Archetype governance covers multiple activities. The following tasks are part of the archetype governance process:

- **Establishment of governance style.** Each organization, initiative or project developing or using archetypes must select a decision-making mechanism and a coordination mechanism between the participants. Section 5.3 describes alternative governance styles.
- **Assignment of responsibilities.** The assignment of responsibilities is greatly influenced by the governance style. Several questions should be answered: Who can propose a new archetype? Who can edit existing archetypes? Who can mark an archetype as ready for publication? Who can withdraw an archetype? Section 5.2 describes the composition of a governance board and how to coordinate the participants in the archetype definition process.
- **Definition of the scope.** Archetype governance can be limited to the creation of a set of archetypes for a specific project or domain. However, it is common that archetypes are managed at a regional, national, or international level. This implies that governance has to define the scope of developed archetypes and coordinate them. This is especially important since archetypes, by their nature, are designed for reuse among different domains. Archetype governance has to avoid the overlap of archetypes and maximize their genericity. Section 5.4 describes different strategies to guide archetype development.
- **Organization of development priorities.** Semantic interoperability projects are usually constrained by personnel, financial and time limitations. In those cases, archetype governance has to prioritize which archetypes should be defined first in order to cover the essential health information needed by the project. This decision can be influenced by current health care trends and needs,

by an evaluation of the healthcare and socio-economic impact of having those archetypes available, or just by the expertise or area of knowledge of archetype editors. An example of this is the openEHR COVID-19 Project [79] started in March 2020 as a community effort to model COVID-19 screening and reports. This initiative was considered a priority, and succeeded in developing a complete set of COVID-19 archetypes in the period of three weeks.

- **Provision of design guidelines.** As explained in Section 4.4, a design guideline describes the general rules, best practices, and common agreements to ensure a consistent development of archetypes at a national, regional or organizational level. It is a responsibility of archetype governance to develop and to provide these archetype guidelines to archetype editors.
- **Quality assurance.** The archetype modeling and development process is the main responsible to ensure the individual quality of the resulting information models. However, archetype governance also influences in the quality of archetypes in a wider perspective. A set of multiple archetypes have to maintain a coherence and balance in their information structures, the overlap with already existing archetypes must be minimized, archetype reuse should be maximized, and the final archetypes have to be fit for purpose and fulfill the initial information needs. Archetype governance has to take care of all these aspects.
- **Publication of archetypes.** Archetype governance should provide the means to publish archetypes and to make them accessible by the potential users. The publication of archetypes can be done using a simple web page or rely on advanced management systems. More information about existing governance information systems can be found in Section 3.2. Additionally, Section 6.1.2 describes LinkEHR Model Manager in detail, a CIM governance system that is a practical result of this thesis.
- **Coordination with terminology developments.** Archetype development has to be synchronized with terminology developments. Terminologies are in constant evolution to maintain them close to health discoveries and techniques, for example adding concepts for new diseases, drugs or procedures. Those con-

cepts are used to semantically describe the structure and contents of health information in combination with archetypes. It is necessary that archetype governance is aware of the evolution of terminologies, decides which version has to be used, and reviews older archetypes to confirm their validity. Archetype governance can be also in charge of providing and maintaining terminology value sets used by archetypes, although this role is usually a responsibility of terminology servers.

- **Coordination with other archetype governance initiatives.** Reaching a global consensus about the structured content of archetypes requires a collaborative effort of multiple countries and organizations. The principles of sharing and reuse of archetypes must be incorporated in the core of archetype development, as the only means to make possible a common formalization of health information. This collaboration is extensible not only to archetype development communities, but to any CIM development effort. CIMI [19] is an example of those efforts to define common clinical information models that are independent of a particular standard or technology. Archetype governance will be responsible of promoting a good understanding and collaborative work among different modeling initiatives.

5.2. Management of participants and roles

Every collective effort needs guidance to maintain the focus on the objectives and to organize the work to be done. In archetype governance, a management board must exist to assume this responsibility.

Definition 2. *The **archetype governance management board** comprises individuals or organizations with the authority needed to define the objectives, priorities, strategies, organization of the participants, editorial policies and guidelines of the governance activity.*

Its members can be selected by decision of a superior entity, by election, or merits. Ideally, the members of the management board should be experienced archetype editors, since some of their decisions have a direct impact on how the archetypes will be developed. They also act in representation of the stakeholders. In our context, the stakeholders are the potential users of the archetypes.

The management board can undertake editorial activities, i.e. participate in the definition process of the archetypes, or delegate it in other people. However, it is necessary that those roles are not confused, even if assumed by the same people. The editorial activities are those actions aimed to the creation or modification of archetypes. These are the activities defined by the AMM in Chapter 4.

The responsibility of the management board is to select the group leader of each project or archetype development work group. The group leader can designate the other members of the work group, including clinical experts, terminology experts, technical experts, and a multidisciplinary clinical support team.

Independently of the background of each participant of a work group, they can be classified in four basic roles that are orthogonal to their professional background. Those roles are:

- **Editor.** A person that has the right to modify the archetypes, including their description, keywords, changes in its lifecycle status, creation of new versions, modification of the archetype structure, modification of the terminology bindings, etc. Participants with an editor role are the main responsible of the archetype definition and evolution.
- **Contributor.** A person that can provide information about the clinical domain of the archetype, or about the use of the underlying RM. A contributor can review the information structure and terminology of an archetype, provide comments, suggest modifications, or raise concerns.
- **Validator.** A person that revises and validates the final structure and contents of an archetype, to check if it fulfills all the requirements of the potential users. A validator provides the necessary quality assurance to publish an archetype

safely, indicating that it is ready for use. Otherwise, a validator can make comments or require a revision and modification of the archetype.

- **Translator.** A special type of contributor whose expertise is language translation. It is recommendable to postpone the natural language translation of an archetype until a stable version of it exists. This provides a greater coherence in the translated terms. Translators must be proficient in both the original language of the archetype and the translation language. It is recommendable that professional translators and experts in the medical language do this work.

5.3. Governance style

Governance can adopt different styles of decision-making and control of the activities of the governance process. According to the characteristics of the management board, the organization responsible of the governance, the type of project, or the scope of the developed archetypes can also influence the governance style. We define governance style as follows:

Definition 3. Governance style is the combination of characteristics that define how the management board makes decisions and conducts the development and management of archetypes.

In this thesis we identify four types of governance styles and five governance characteristics that are transversal to them. They are listed in Table 5 and described in depth later.

Governance styles	Governance characteristics
Authoritarianism	Proactive
Democracy	Agile
Meritocracy	Adaptable
Do-ocracy	Conciliatory
	Traceable

Table 5. Types of governance styles and applicable characteristics

The governance style defines who and how decisions are made during the governance process. There are four possible governance styles, mostly derived from existing political systems, but also derived from organizational structures of collaborative development initiatives.

- **Authoritarianism.** The promoters of the semantic interoperability project, or their representatives, control the governance. The requirements and interests of the promoters guide the governance decisions, even over clinical recommendations. Alternative or opposite opinions can exist, but they have little influence in the result. This is the case of initiatives where a committee decides the contents of the clinical models, usually limited by tight release calendars or strict design rules. This style of management is only useful in small projects where the objective is to get a quick solution, but there is no requirement for real semantic interoperability or harmonization of the models with other modeling initiatives.
- **Democracy.** Every member of the editorial and governance process has a say in what it is done. A qualified majority will take the decisions and decide the planning and acceptance of the work. A clear benefit of this approach is that everybody has the opportunity to express their opinions and to defend them. The main drawback is that the decision-making process may be slow.
- **Meritocracy.** In a meritocracy, the most qualified people take the decisions. In the case of archetype modeling, the group should include medical experts in the particular domain of the archetype and technical experts in the EHR systems and in the selected RM. This governance style usually ensures the highest quality of the developed archetypes, although it can also fall in the error of not incorporating relevant comments of other members of the community.
- **Do-ocracy.** In a do-ocracy, whoever does the job gets it. Everyone does jobs that they think that are needed, no matter how well they are qualified and without everyone's input. This is a common approach in open communities of users or open source projects. The main handicap of this approach is that the developments are not centrally prioritized, or if they are, it may be that nobody

works on them. This approach is followed in the Norwegian national openEHR CKM [80].

Archetype governance experiences indicate that it is common to follow a combination of the previous styles, in order to take advantage of the best characteristics of each of them. For example, having members designated by the political organizations together with expert members selected by the industry, or having an open community of developers (do-ocracy) guided by prominent members (meritocracy).

There are additional and transversal characteristics that are intrinsic to a successful archetype governance, whatever it is the governance style.

- Governance has to be **proactive**. Archetypes should be aligned to current clinical information needs and clinical knowledge. Archetype governance should not wait for someone to ask for new or updated model. It should be aware of the changes of requirements and promote the renewal of archetypes whenever it is needed. This is especially important in the maintenance of terminology bindings, since terminologies usually have a faster evolution speed than information structures, which are more stable in time. Terminology bindings should be frequently revised and updated to incorporate the changes in the terminologies.
- Governance must be **agile**. It must react quickly to the changes and new needs in healthcare. For example, to provide clinical information models in the case of a new illness or a global pandemic. However, agile is not the same as uncontrolled. The processes for creating and validating new archetypes have to be maintained to ensure their final quality.
- Governance must be **adaptable**. Not all interoperability projects have the same requirements. It is not comparable the effort required to create a nation-wide semantic interoperability infrastructure based on archetypes, to the effort required in the context of a single hospital. Although the basic characteristics of archetype governance should be maintained, it is recommendable to adapt the

processes and strategies to the context of development and to the available resources.

- Governance must be **conciliatory**. Even in the case of an authoritarian governance, the path to success in semantic interoperability relies on conciliation and agreement between multiple preferences and requirements. Archetype governance must satisfy the needs of multiple potential users, or at least it has to encourage dialog and agreement among them. Imposing the contents of archetypes from a single point of view might work only in strictly controlled scenarios (for example, when a national law or standard already mandates which clinical contents must be shared). However, open communities and international semantic interoperability initiatives might not accept these imposed models since they do not come from a broader consensus.
- Governance must be **traceable**. There must be a registry of all the changes produced in the archetypes, including the identification of the authors of those changes, the justification for making them, and, in general, it must keep any other audit information needed to have an uninterrupted view on the evolution of the archetypes. It must be possible to access and retrieve all of the historic versions of the archetypes in order to provide support to existing systems that have not been updated to use newer definitions.

5.4. Archetype development strategy

Semantic interoperability projects usually require the definition or reuse of multiple archetypes. Those archetypes will probably cover different areas of information, i.e. they represent information structures for different health domains. It is necessary to plan and organize the order in which the archetypes will be developed. It is a responsibility of archetype governance to decide which strategy to follow to design and develop archetypes.

We propose two different archetype development strategies, according to how the development is organized. The archetypes obtained will be probably the

same or equivalent at the end of either strategy, but each of them has its own benefits and problems.

- **Development in width.** In this strategy, the efforts are focused on identifying and creating a first version of the archetypes required to cover the information needs of the semantic interoperability project.

The main benefit of this approach is that we can determine the main relationships between different archetypes and therefore design an information model architecture where the coverage of each archetype is established since the beginning. It will be possible to design archetypes that are more generic and reusable because we know the different contexts where they will be used.

The disadvantages of a development in width are that we cannot divide the development task in specific or specialized work groups. On the contrary, this strategy requires a continuous collaboration between all the members of the group. Another concern is to assume that it is possible to know in advance the complete extent of the work to be done, which may not be possible in complex semantic interoperability scenarios.

- **Development in depth.** In case a very specific semantic interoperability project, or in case it is required to integrate national or regional requirements into the archetypes, it is possible to develop archetypes in depth. The objective is to model and implement very detailed archetypes at once, all the way down to the last element of information.

When archetypes are developed in depth, the work to be done includes creating templates, which are profiles of the archetypes to adapt them to specific requirements. It also includes the translation of archetypes to different languages, to support multilingual requirements or to use localized terms.

In order to facilitate the governance of very specific profiles of archetypes, it is possible to create namespaces, i.e. identifiers of sets of archetypes that are closely related. The use of namespaces allows a better control of the provenance of the archetypes, and an improved control over the publication of the complete set of related archetypes.

The benefits of development in depth is that archetype modelers can be grouped by areas of specialty and focused on modeling all the requirements of a specific knowledge domain.

Another benefit is that it is possible to complete a set of archetypes that cover the information needs of a domain and start using them without needing to wait to the completion of all the other archetypes.

The disadvantages of a development in depth arise if, once the development has started, it is found that some additional requirements do not fit in the developed archetypes. This would require an adaptation of the work done, or in the worst case, it would need to start over the modeling process.

The two development strategies described are compatible. It is reasonable to combine them to obtain better results. In a big national or international project, it would make sense to start with a development in width, to identify the main domains or groups of archetypes, and then create different work group to model the details of each domain in depth. Archetype governance should decide which strategy to follow or when it is recommended to combine them.

5.5. Quality assurance

Quality assurance of archetypes must be present during the modeling, development, and governance activities. Alberto Moreno defines three different models for evaluating the quality of CIMs [81], which also apply to archetypes:

- **Development quality model.** It includes the human and organizational factors that are required for defining, drafting, reviewing and publishing CIMs.
- **Product quality model.** The definition of objective requirements for a CIM, which could allow a potential adopter to determine what level of semantic interoperability it may support.
- **Quality in use.** It includes specific measurements for usability of the models.

Following these definitions, the development quality model corresponds to the validation phase of AMM (see Section 4.5.4.). This validation phase includes

identifying errors, inconsistencies, absences of information, or misleading specifications before publishing the archetypes.

In the context of archetype governance, the product quality model determines the requirements to verify the overall quality of archetypes in relation to the overlap with other archetypes, the consistency of terminology bindings, and the fulfillment of the initial requirements for the archetypes. A more detailed description of these requirements is provided hereunder.

Once the archetype modeling work group defines a set of archetypes, and they confirm their technical and semantic validity, it is responsibility of the archetype governance to validate them in the context of already existing archetypes and CIMS. The overlap with other archetypes has to be examined. A complete overlap would be a reason to reject the new archetype definitions, although in this case, the archetype modeling work group should have previously detected and addressed the problem during the analysis and validation phases of archetype. However, it is possible that a partial overlap is detected and that it does not invalidate the new archetypes. In this case, archetype governance can either initiate a refactoring process of the archetypes or improve their scope description. In the first case, refactoring means identifying common information structures used by different archetypes and create new independent archetypes to be reused. In the second case, it is necessary to clearly describe the scope, recommended use and misuse of the archetypes. Making clear the valid use cases for each archetype is a requirement to avoid misunderstanding in the potential users.

A similar revision must be done with regard to terminologies and ontologies. Archetypes are bound to terminology value sets that represent valid codes for coded elements in the archetype. These value sets can be used by multiple archetypes. For example, a value set representing the valid SNOMED CT codes for gender. Archetype governance has to review that the value sets are reused whenever it is possible and avoid the existence of overlapping value sets.

Finally, but not less important, archetype governance has to verify that the resulting archetypes fulfill the initial requirements of information interoperability.

These requirements are present during all the archetype modeling and development process, but it may happen that the modeling work group loses the focus on the global picture of needs and concentrates in resolving their specific requirements. Thus, the results might not maintain a design coherence with other archetypes. To maintain the coherence of this global view is a responsibility of archetype governance.

5.6. Publication and visibility

Information systems cannot be semantically interoperable without sharing clinical information models (archetypes in our context). Archetypes must be published and made available somewhere on either the Internet or the private network of the organizations, so that they can be retrieved when needed.

Once the archetype modeling and development has finished, the last phase (Phase 5 of AMM, see Section 4.5.5) is the publication. At this point, the archetype modeling work group and the management board have to collaborate and agree if the archetype is ready for use in production. This depends on the coverage of the information requirements for the archetype and its congruence with other existing archetypes. If the answer is affirmative, the archetype is published in the governance platform.

A second decision that the management board has to make is about the visibility of the archetypes. That is, to decide if the archetypes are publicly available, allowing that any person use them, or if they are kept private inside the organization or working group. The visibility of archetypes can be decided since the very first stages of the development, making public even the early drafts of the archetypes and accepting public comments about them through the complete modeling process.

The public visibility of archetypes is a basic principle of a successful semantic interoperability. Without it, archetype modelers may deviate from the information requirements of the potential users. Moreover, the communication and processing

of data between different systems will be extremely limited, since they cannot access to the archetype definition. However, some use cases can require the limitation of the visibility of archetypes. This would be the case of working in a very specific integration project, or if any license limitations apply.

An option is to adopt a mixed approach. That means to work with private archetypes during the first stages of development and make them open for public participation later, in the final revision and validation phases. However, this approach harms the quality of the resulting archetypes, and can delay the archetype development process, since concern about the core archetype scope and structure might not be detected until it is too late to incorporate the required changes, or require starting over the modeling process.

5.7. Alignment of archetype governance activities to PRINCE2 project management processes

This chapter presents a collection of topics, activities, and responsibilities that any archetype governance process should incorporate. They are presented as good practices and recommendations, and not as a strict methodology for archetype governance. These recommendations can be summarized in the following activities:

- Select a management board, in charge of organizing the archetype modeling work, defining the strategies, editorial policies and guidelines, and responsible of keeping the development focused on the initial requirements.
- Adopt a governance style, deciding about the degree of empowerment of the participants in the decision-making process.
- Promote collaboration and reuse of clinical information models, with the objective of creating reusable archetypes and minimizing overlap and incompatibilities between different modeling initiatives.
- Promote coordination with ontology and terminology developments.

- Control the quality of the results, by establishing the appropriate mechanisms to validate the contents of the archetypes prior to their final publication. These mechanisms should include the ability of reviewing and incorporating changes suggested by the community of users.
- Publish the results, making archetypes available for semantic interoperability projects and to strengthen the reuse of the models.

Archetype governance is not essentially different from other governance activities inside an organizational or IT development project. It has its own particularities, which are the ones described in this chapter, but they can be incorporated to existing and well-established project management methodologies. The main two internationally recognized project management methodologies are Project Management Professional (PMP) [82] and PRjects IN Controlled Environments (PRINCE2) [83]. We will use PRINCE2 as an example of how to combine the methodology with the recommendations presented in this chapter. Using any other management methodology, such as PMP, would require a similar correspondence analysis.

PRINCE2 is a project management methodology and certification programme. The objective of PRINCE2 is to provide the generic guidelines for project planning, responsibilities delegation, monitoring and control. One of the principles of PRINCE2 is its ability to be adapted to the needs of any specific project. Therefore, archetype governance can use PRINCE2 as an underlying management methodology by aligning the archetype governance recommendations to the PRINCE2 processes.

There are multiple roles played by the participants in a PRINCE2 project, but two of them are especially relevant in our context of archetype governance:

- **Project board.** The members of the board provide a unified direction to the project and are accountable for the success or failure of the project. This is a similar role to the one played by the archetype governance management board, which is in charge of controlling the overall governance.

- **Project manager.** A project manager runs the project on behalf of the Project Board and manages the project on a day-to-day basis. It would be a similar role to the archetype modeling work group leader, who is in charge of guiding a particular domain or subset of archetypes.

PRINCE 2 divides the projects into stages. At least, there is an initial stage dedicated to the planning of the project, followed by as many development stages as needed to complete the project objectives. PRINCE2 also describes seven processes, which define the different activities performed inside a project. These processes are:

- **Starting up a project.** The objective of this process is to define the reason for the project (business case) and to make sure that the project is viable before initiating it.
- **Directing a project.** The objective of this process is to provide the continuous management of the project, including the authorization for the project initiation and closing, and to provide direction and control during the execution of the project stages.
- **Initiating a project.** The objective of this process is planning the execution of the project, including the development of strategies, guidelines and documentation of the project.
- **Controlling a stage.** The objective of this process is to monitor that the development of each project stage stays within a predetermined tolerance.
- **Managing product delivery.** The objective of this process is to guarantee the delivery of the products of the project at each stage.
- **Managing stage boundaries.** The objective of this process is to review the success of the current stage, to approve the initiation of the next stage, and to review and update the project plan if it is necessary.
- **Closing a project.** The objective of this process is to check that the project objectives are completed, transfer the results to the client and identify lessons learned during the execution of the project.

Figure 18 represents the workflow and the relationships of the described processes. At the initial stage, starting up a project confirms the viability of the project, or cancels it in case it is not feasible to achieve the expected results. If the project is approved, the directing a project process initiates the project, which means creating the detailed execution plan. Then, an iteration of development stages take place, each of them with their own control and execution activities, including the validation of the partial results obtained at the end of each phase. Finally, once all development stages finish, the closing a project process confirm the complete results of the project and delivers it to the client.

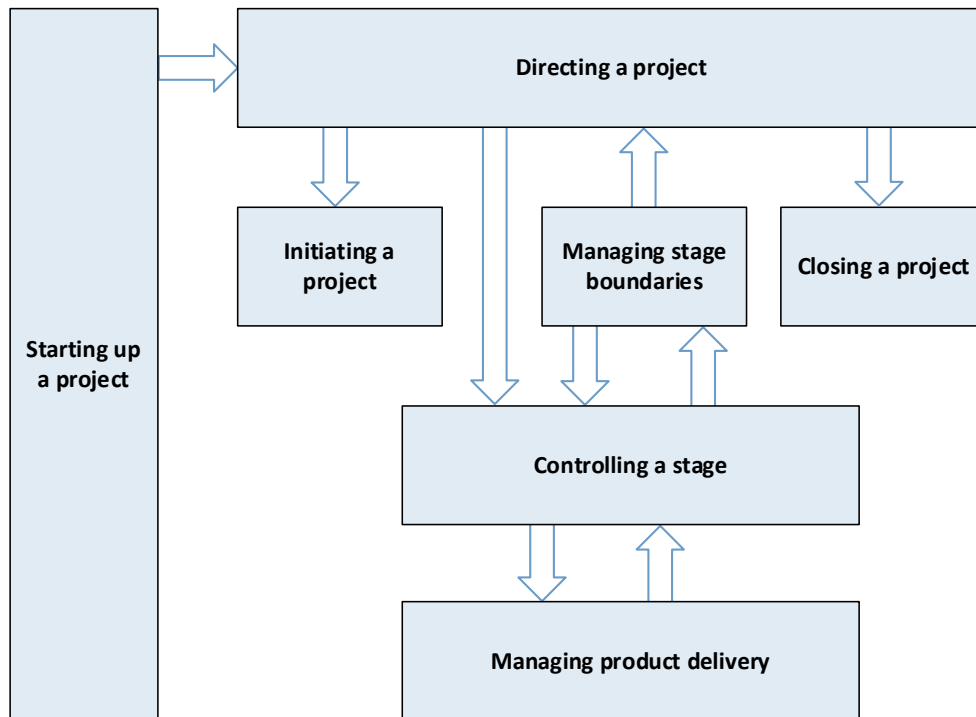


Figure 18. Workflow of PRINCE2 processes

Our premise is to consider any semantic interoperability development project as a PRINCE2 management project. Based on this assumption, Table 6 shows

the correspondence between PRINCE2 processes and the main archetype governance activities. It also identifies where the actual archetype modeling phases take place.

PRINCE2 process	Archetype governance activities
Starting up a project	<ul style="list-style-type: none"> ▪ Definition of the scope
Directing a project	<ul style="list-style-type: none"> ▪ Selection of the archetype modeling work group leader ▪ Establishment of governance style
Initiating a project	<ul style="list-style-type: none"> ▪ Assignment of responsibilities ▪ Organization of development priorities ▪ Provision of design guidelines
Controlling a stage	<ul style="list-style-type: none"> ▪ Execution of the archetype validation phase
Managing product delivery	<ul style="list-style-type: none"> ▪ Execution of the archetype modeling phases
Managing stage boundaries	<ul style="list-style-type: none"> ▪ Quality assurance
Closing a project	<ul style="list-style-type: none"> ▪ Publication of archetypes

Table 6. Correspondence between PRINCE2 processes and governance activities

The scope of the interoperability project delimits the scope of the archetypes that will be developed to cover the information needs of the project. As the first stage, we need to check that it is possible to develop the archetypes with the existing resources and knowledge. This fits in the starting up a project activity of PRINCE2.

If the project is viable, the governance management board nominates a leader of the archetype modeling work group, and selects a governance style. This decision defines who will manage the project and how it is controlled. This is also

the aim of the directing a project process, where a project board delegates in a project manager to deal with the day-to-day development tasks.

The leader of the archetype modeling group selects the other members of the work group, organizes the specific work and defines the modeling guidelines. These are all tasks identified as part of the PRINCE2 initiating a project activity, where the project plan is developed, containing all the details that guide the successful execution of the project.

Archetype quality assurance is the activity that confirms the validity of the developed archetypes. In particular, it confirms the conformity to development guidelines and to the initial requirements. This is achieved in two different moments. The development quality is evaluated during the validation phase of AMM and confirms that the actual archetype development is done correctly. It is aligned to the controlling a stage process of PRINCE2. Additionally, the verification of the quality of the complete set of archetypes, to check that they are conformant to the initial quality requirements, is aligned to the managing stage boundaries process.

The contact point between archetype governance and the archetype modeling phases described in the AMM takes place in the managing product delivery activity. This activity is where a work group leader performs its role guiding the work of other archetype modeling and development participants.

At the end of the archetype modeling activities, the archetypes are the results of the project. They have to be provided to the client (the users of the archetypes). In PRINCE2, the closing a project activity is also in charge of the final acceptance of the results and the deployment to the client.

Chapter 6

Software development and use cases

This chapter describes the practical results of several software developments made to support the modeling and governance of archetypes. The author of this thesis had a lead role in the design and development of these implementations, in collaboration with other members of the IBIME group at the UPV and in VeraTech for Health.

This section also describes the most relevant projects where the author has participated and provided guidance on the application of the archetype modeling methodology and the archetype governance principles.

6.1. Software developments

6.1.1. LinkEHR Studio

In order to apply the guidelines of the AMM it is necessary to use specialized tools that help and guide archetype authors during the archetype definition process. Archetype editors are tools that facilitate the definition of data structures, data constraints and terminology bindings. Although archetypes are expressed in the Archetype Definition Language (ADL), a textual format that can be complex to edit manually (Figure 19), visual archetype editors are essential to facilitate the work of clinical users in charge of defining archetypes. Clinical users provide the domain knowledge to be represented or supported by the archetype information structure, but they might not know all the technical details of the underlying RM. It is a responsibility of the tools to hide the complexity of the RM and let the users focus on expressing the clinical constraints and data structures of the archetype.

Archetype editors should also facilitate the definition of terminology mappings by connecting to a terminology server. While terminology servers store the complete terminologies and value set definitions, an archetype editor should provide a seamless access to them. This access is necessary to complete the archetype definition by adding the semantic mappings and the value mappings. Since one of the initial steps of the archetype design process is to name the different elements or nodes of the archetype structure, it will be useful that terminology servers provide natural language-based search capabilities that help in the automatic discovery of terminology mappings.

```
OBSERVATION[at0000] matches { -- Blood Pressure
  data existence matches {1..1} matches {
    HISTORY[at0001] occurrences matches {0..1} matches { -- history
      events existence matches {0..1} cardinality matches {1..*; unordered} matches {
        EVENT[at0006] occurrences matches {0..*} matches { -- any event
          data existence matches {1..1} matches {
            ITEM_TREE[at0003] occurrences matches {0..1} matches { -- blood pressure
              items existence matches {1..1} cardinality matches {0..*; unordered} matches {
                ELEMENT[at0004] occurrences matches {0..1} matches { -- Systolic
                  value existence matches {1..1} matches {
                    C_DV_QUANTITY <
                      property = <[openehr::125]>
                      list = <
                        ["1"] = <
                          units = <"mm[Hg]">
                          magnitude = <|0.0..<1000.0|>
                          precision = <|0|>
                        >
                      >
                    >
                  >
                >
              >
            >
          >
        >
      >
    >
  >
}
ELEMENT[at0005] occurrences matches {0..1} matches { -- Diastolic
  value existence matches {1..1} matches {
    C_DV_QUANTITY <
      property = <[openehr::125]>
      list = <
        ["1"] = <
          units = <"mm[Hg]">
          magnitude = <|0.0..<1000.0|>
          precision = <|0|>
        >
      >
    >
  >
}
}
```

Figure 19. Example of an archetype expressed in ADL

Another functionality of an archetype editor is to enable methods for the validation of the archetypes. This is a threefold validation:

- **Structural validation.** Archetypes must follow the structural rules imposed by an underlying RM. Archetypes can constrain the generic structures of the RM, but they cannot create new data structures that are not supported by the RM.

A structural validation focuses on confirming that the archetypes are congruent to the RM structures. This also applies to specialized archetypes in relation to their parent archetypes.

- **Semantic validation.** Terminology mappings provide the specific semantics of archetype information structures. It is necessary to validate that the semantic mapping of the information structure is consistent with value mappings, and that the overall semantics of the archetype is congruent. For example, an archetype element mapped to the SNOMED CT concept 4241000179101 [Laboratory report (record artifact)] may contain an information element for the type of laboratory test whose values should be descendant concepts of 15220000 [Laboratory test (procedure)].
- **Functional validation.** It is necessary to validate that archetypes are useful in the real world for recording or querying clinical data. A wider set of potential clinical users of the archetypes, different from those who defined them, should review the information structure defined and its constraints. They should focus on the clinical contents of the archetypes, and not in their technical definition. A mechanism to achieve this by generating simulated screen forms, or other alternative visual representations of the archetypes such as mind maps.

Finally, to facilitate the coordination between people creating archetypes, and to promote the reuse of existing archetypes, archetype editors should work together with an archetype governance platform. The functionality provided by this platform includes the management of the lifecycle of the archetypes, the archive of previous versions, the publication of archetypes to make them accessible by other users or information systems, the provision of search capabilities to locate archetypes, and the management of related groups of archetypes used together in specific scenarios.

There are few archetype editors available, and even fewer that comply with the previous requirements. Moreover, most of the existing editors focus on the openEHR RM only. This limits their applicability to other RMs.

The openEHR Archetype Editor [71] is the reference tool provided by the openEHR Foundation to define openEHR archetypes. It provides a clinical-oriented interface that eases the definition process of archetypes by hiding most of the RM complexities related to the representation of context information or other data structures that are not relevant for clinical modeling. Currently this tool only supports openEHR, being its interface highly coupled to this specific RM (Figure 20). It allows defining terminology mappings locally, and it is connected to a terminology server, although it is a closed solution by the company Ocean Health Systems whose contents cannot be updated. It is also connected to the openEHR Clinical Knowledge Manager (CKM) for searching archetypes and downloading them to the editor, but it does not support a write back to the CKM. Finally, it allows visualizing the edited archetype as a simulated form or as a readable summary table and can export it in different formats such as XML or OWL.

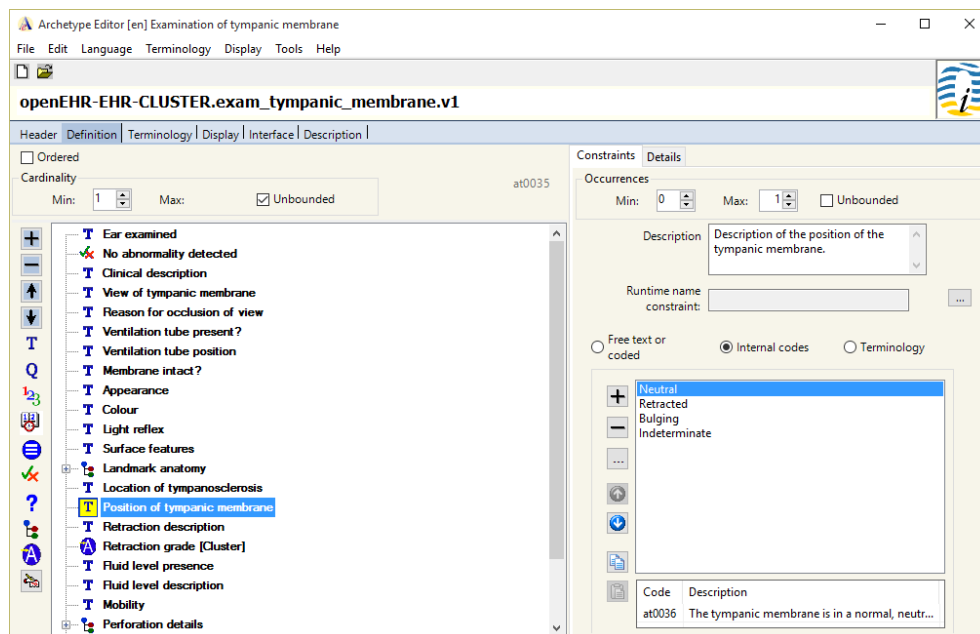


Figure 20. OpenEHR Archetype Editor

A companion tool for the openEHR Archetype Editor is the Template Designer (Figure 21). A template, according to openEHR, further constrains a set of archetypes to adapt them to local scenarios of use. The functionality supported by

the Template Designer is similar to the one included in the Archetype Editor. It also includes the connection to the Ocean terminology services, and the ability of generating simulated forms. However, it is not connected to the openEHR CKM and can only interact with a local folder as a repository of archetypes and templates.

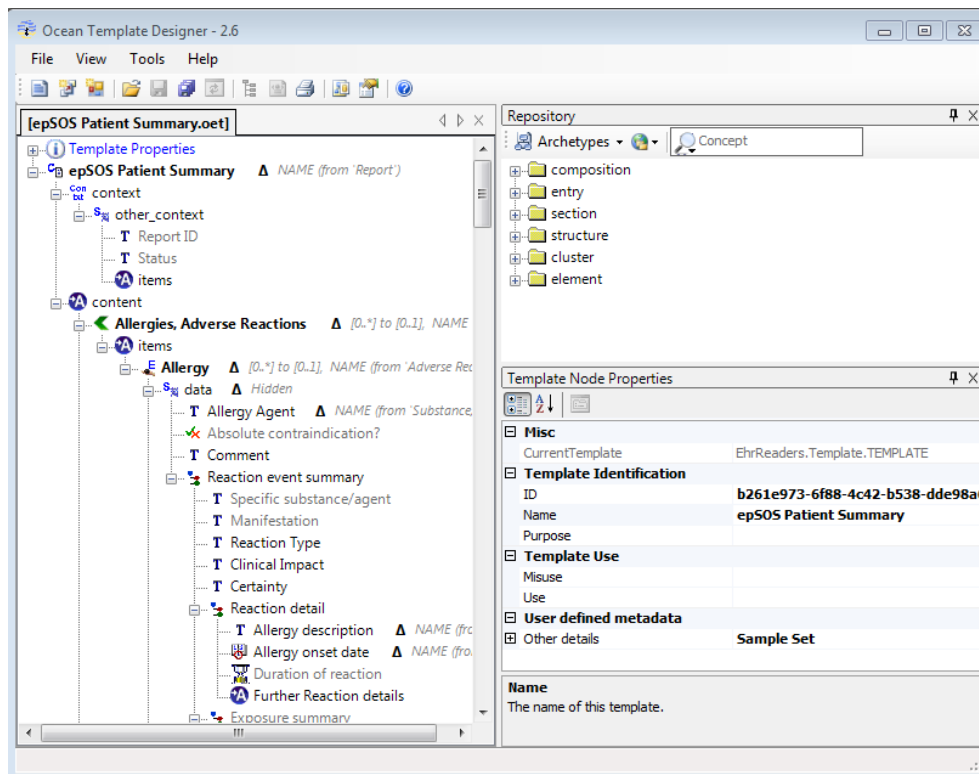


Figure 21. OpenEHR Template Designer

An evolution of both the openEHR Archetype Editor and the Template Designer is the Archetype Designer [54] (Figure 22), a web-based tool allows creating archetypes and templates. It is based on ADL 2, that serves for both archetype and template definition. Developed as open source, it provides a basic support for archetype and template edition, but it lacks most of other advanced functionalities. It does not provide any additional visualization format for the archetypes, only the possibility of exporting a mind map. It does not have a connection to any terminology server and to any archetype governance platform either. It does provide the possibility of storing the archetypes and templates in cloud services such as GitHub,

Google Drive, Dropbox, or OneDrive, but that is not a substitute of a complete governance system.

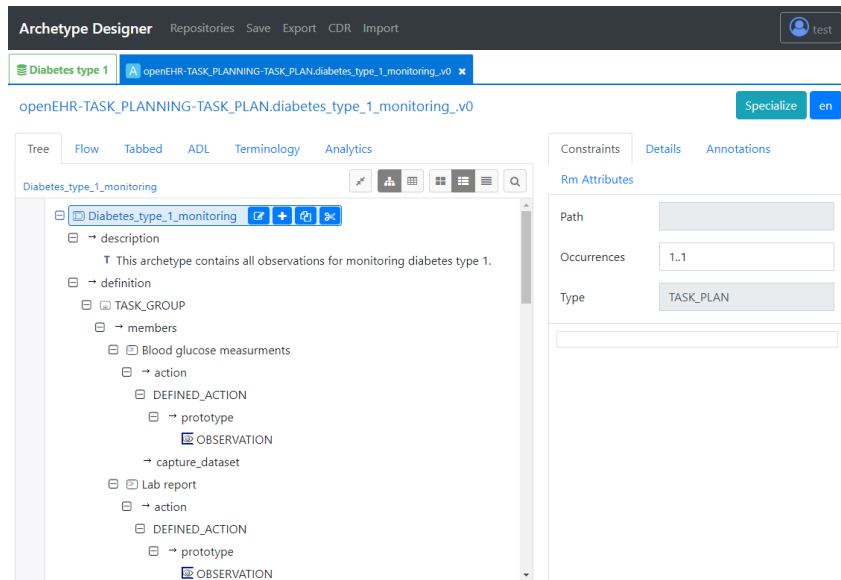


Figure 22. Archetype Designer

Finally, another tool provided by the openEHR Foundation is the ADL Workbench [84]. This tool allows testing new functionalities of the archetype model, especially those related to the new ADL 2 specifications [72]. It does not provide edition capabilities, although this is a planned functionality in the future. The ADL Workbench supports multiple RMs, defined in the Basic Meta-Model format (BMM). BMM is a proprietary format created by openEHR to represent a computable definition of Object-Oriented reference models [85]. The ADL Workbench provides a tree view of the archetypes, showing all the properties the reference model and the constraints defined in the archetype. This tool is oriented to the technical examination of archetypes, rather than their clinical edition or review.

Back in 2005, the available options were more limited. Only the openEHR Archetype Editor and the Template Designer existed. Standards such as ISO 13606 (formerly called EN13606) did not have tools to edit archetypes. As part of our research at the IBIME group of the ITACA institute of the UPV, we needed editing archetypes based on other standards such as HL7 CDA, in order to use them for data

transformation purposes. Our contribution in the field of archetype editors was LinkEHR Studio.

LinkEHR Studio (www.linkehr.com) is an archetype editor that fulfills the needs of both technical and clinical professionals that design archetypes. LinkEHR Studio was first introduced in 2006 [86] and has been in continuous development since then, incorporating some of the results of this thesis. Many researchers and institutions have used it since its first release [41,87–94]. National and regional health services have also used LinkEHR Studio, for example, the National Health Services England, the Ministry of Health of Uruguay, the Ministry of Health of Spain, the Valencia Regional Health Service in Spain, or the Minas Gerais Regional Health Department in Brazil.

A detailed description of the internal architecture, data model, and type system of LinkEHR Studio (formerly called LinkEHR-Ed) is available in [69]. We will provide a short summary of how it works here.

LinkEHR Studio is a visual tool that allows the edition of archetypes based on different RMs, and the generation of data transformation programs that translate non-normalized XML data into standard XML documents (Figure 23). In this description, we will focus on the archetype edition capabilities. Data transformation and normalization is an essential step towards semantic interoperability of existing EHR systems, but it is out of scope of this thesis.

LinkEHR Studio facilitates the definition, review, and validation of archetypes, and has three main principles:

- **It is independent of any particular RM.** We can import new RMs into LinkEHR Studio, and we can create archetypes by constraining any of those RMs.
- **It is usable by both technical and clinical professionals.** LinkEHR Studio offers different visual interfaces for technical or clinical professionals. Archetypes are represented as a tree that can be configured to show the complete archetype structure, including all the classes and attributes of the RM, or be

configured to show only those elements that are clinically relevant for constraining.

- **It provides advanced functionalities based on the formal nature of archetypes.** LinkEHR Studio enables the automatic generation of sets of reference materials from clinical archetypes. This includes alternative representations of the clinical information model, validation programs, simulated formularies, implementation guides and sample data instances. Most of these processes are described in detail in [95].

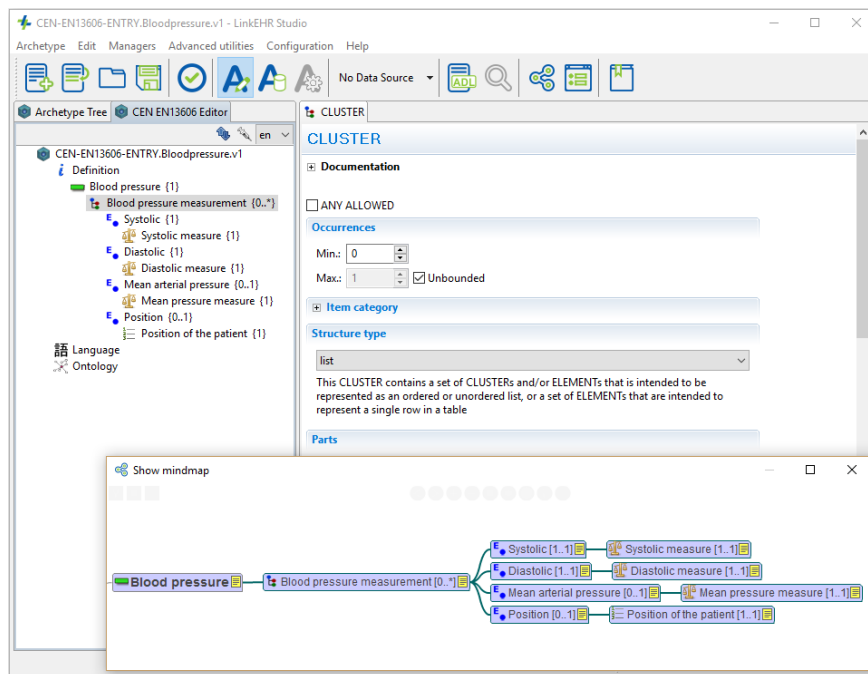


Figure 23. LinkEHR Studio

In the following paragraphs, we describe the contributions of LinkEHR Studio that support different phases of the AMM and archetype governance activities.

a) Multiple reference model support

LinkEHR Studio has the capability of importing and using any RM to create archetypes. During this process, the user only has to select which classes will be available as root of new archetypes (Figure 24).

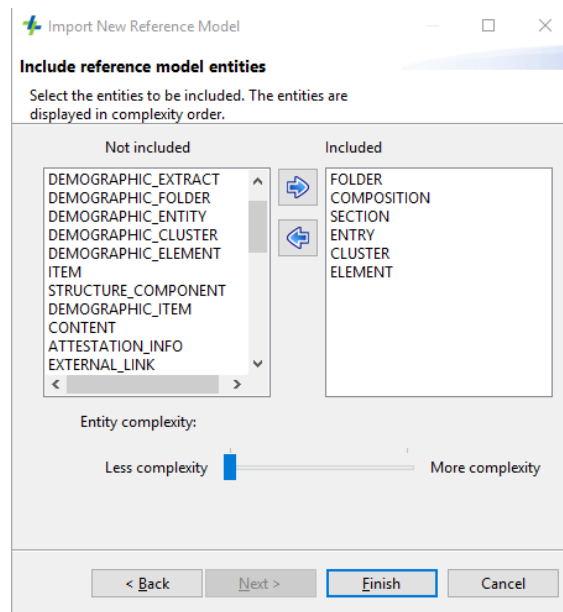


Figure 24. Import of a new RM in LinkEHR Studio

Several reference models have been already used, including ISO 13606, openEHR, HL7 CDA [96], HL7 FHIR [97], ASTM CCR [98], HL7 CCD [99], Clinical Information Model Initiative RM [19], CDISC ODM [100], and MedXML MML [101].

LinkEHR Studio was the first editor to introduce a multiple reference model support in 2005. Originally, it only supported RMs defined in XML Schema. This choice fits perfectly with RMs such as HL7 CDA, CDISC ODM or ASTM CCR, which are natively defined by an XML Schema.

LinkEHR Studio was later improved with the ability of importing RMs in Basic Meta Model format (BMM). BMM [85] is a format defined by the openEHR Foundation as a computable definition of information models. OpenEHR and CIMI use BMM for the definition of their RMs.

b) Development of specific reference model archetype editors

To support multiple RMs without the need of specific implementations, LinkEHR Studio represents the archetype structure as a comprehensive and ge-

neric tree. This tree is a succession of classes and attributes, as defined in the underlying RM, starting in a root class and ending in primitive data types (i.e. integer or real numbers, strings, booleans, times and dates). Figure 25 shows an example of the generic archetype tree. In the example, we can see a sequence HL7 CDA classes and attributes that constraint the RM to define an Allergy archetype.

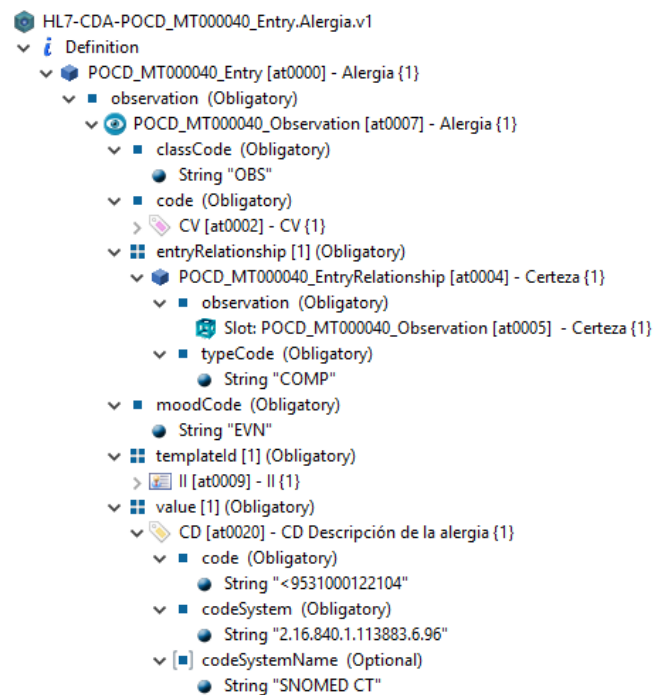


Figure 25. Generic archetype tree representing an Allergy archetype in HL7 CDA

A drawback of the generic archetype tree is that it requires that the users understand the RM in detail. Although building archetypes requires a minimal knowledge of the underlying RM, this cannot be a barrier to incorporate clinical professionals in the archetype development process, as recommended by the AMM. We can hide most of the complexity of the RMs in a basic archetype edition process, but it will require an ad hoc implementation to support each of the RMs.

We define a specific archetype tree as a simplified representation of a RM, including a set of forms to specify constraints over the RM classes and attributes. In the specific archetype tree, we only show nodes that are clinically relevant and hide

the technical attributes of the RM. The decision about which are the clinically relevant nodes is made in a per-standard basis. For example, EHR Standards such as ISO 13606, HL7 CDA or openEHR allow the representation of clinical data but also include attributes and classes for representing technical and context information, such as the unique identifier of the instances, the identifiers of the authors of the information, the versioning, auditing, recording time of the data, etc. There is no need to express any constrain on them in an archetype, since the values will be automatically generated during runtime. Thus, we can hide them in the specific archetype tree. Moreover, we can also hide the attributes used to navigate the archetype structure, as the tool already implements the specific architecture of the RM. Note that hiding does not imply removing. These attributes will be automatically created as part of the archetype definition when saved in ADL, but it will not be shown in the visual interface. Figure 26 shows the same archetype as in Figure 25, but using the specific archetype tree representation. Hiding the technical details of the RM makes it much more understandable and easy to use by clinical professionals.

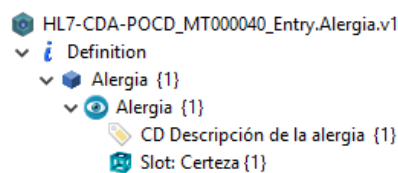


Figure 26. A specific HL7 CDA archetype tree representing an Allergy archetype

Once we have created an archetype information structure in the form of a tree, we have to constraint other properties of the RM, including the valid values of primitive data types. We achieve this by using constraint forms that are rendered when we select a node in the archetype tree. In the generic archetype tree, these forms are also generic, showing only basic constraint options such as the occurrences of classes, the cardinality of container attributes, of the constrain options of primitive data types. In the specific tree visualization, specific forms for each RM are implemented. They allow constraining not only the generic properties, but also the context attributes hidden in the tree. For example, if we define an archetype

called “Spanish patient summary” probably we would need to constrain the territory context attribute to the value “Spain”. This attribute will be accessible from the constraint definition form.

Although we cannot avoid the need of ad hoc implementations to create the specific views, they are based on a generic architecture that can be reused between different RMs. We can configure each imported RM to guide the tool during the interface generation. This configuration is made in the Reference Model Manager section of LinkEHR Studio (Figure 27).

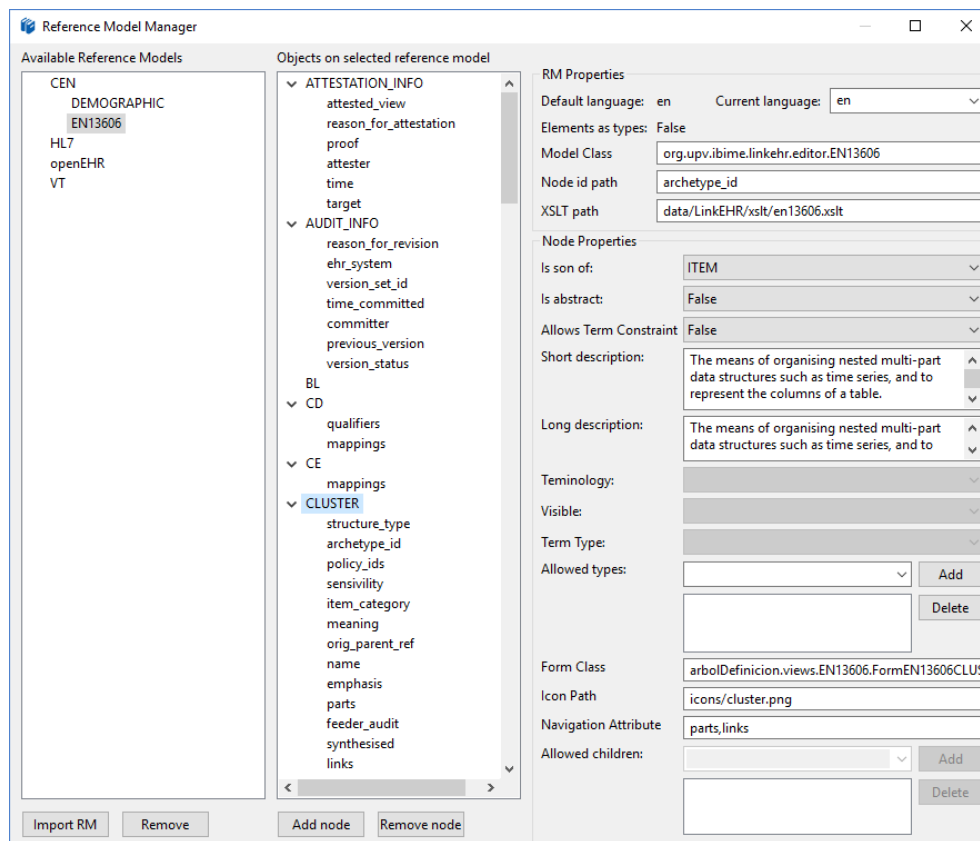


Figure 27. Edition of the documentation of a reference model

For each class of the RM we can specify the main clinical attributes (e.g. the attribute “content” in the case of the ISO 13606 class COMPOSITION). LinkEHR Stu-

dio uses this information to navigate the archetype tree, and hides all other information structures. For each class, the documentation also indicates which is the constrain form associated to it, that will be loaded in the main interface when the class is selected at the archetype tree. The documentation also allows configuring the icons used to represent a class, and writing a short documentation text that describes the meaning of the class, in order to guide the user during the edition process in the form of tooltips. LinkEHR Studio automatically generates a skeleton of this documentation class during the import of a RM. We can edit manually the properties through a configuration interface included in the tool afterwards. We only need to do this configuration once for each RM.

Finally, it is worth to mention that both the specific and the generic views of the archetypes are seamlessly integrated into LinkEHR Studio. The user can rapidly change from the generic interface to the specific one, according to his personal preferences just by clicking a button in the interface.

c) Visualization and validation of archetypes

The development of archetypes requires a continuous analysis by different users coming from different domains to ensure their clinical validity. These clinical professionals should review, discuss and agree an archetype definition that covers all or most of their particular needs. Since many of them may not be familiar with archetypes, or ignore the details of the selected RM, we need to offer different kinds of views of the archetype definition that are more readable or accessible for them. We have developed an automatic generation of four different views based on the formal archetype definition. We will illustrate the different types of visualizations taking the medication use archetype (Figure 28) as an example.

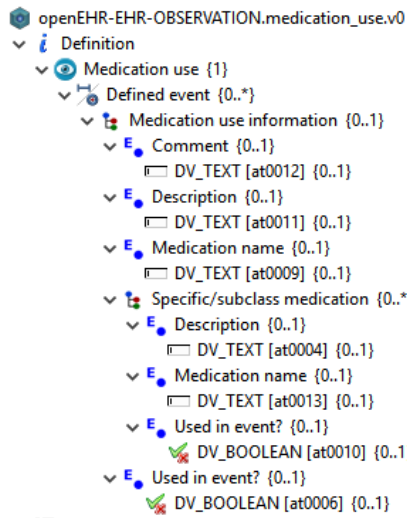


Figure 28. openEHR medication use archetype

- Simulated data-entry formulary.** We can generate HTML forms from an archetype specification that simulate an archetype-based data input screen of an EHR system (Figure 29). The aim of these forms is to evaluate if the archetype accommodates the documentation needs of clinical users. This is a model-dependent visualization that requires a particular development in the tool.

Figure 29. Archetype-based formulary

- Mind map.** Mind maps are an alternative to show the contents defined in the archetype in a visual way (Figure 30). The mind map generation is completely generic, since it represents a graph where the classes of the reference model are the nodes of the graph, linked by the attributes as the arcs. We can generate this representation without needing specific developments for the particular RM used, although it can also take profit of the RM documentation if it exists, for example, using the configured icons or showing the documentation texts as tooltips. We can also export this mind map to edit it in existing mind map software.

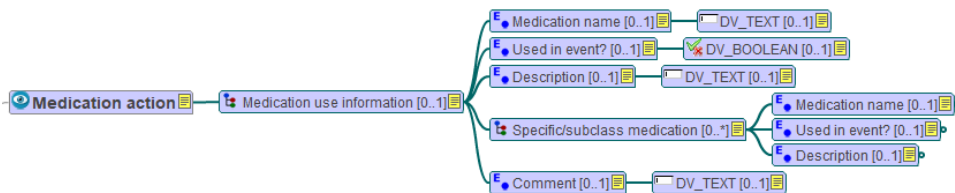


Figure 30. Archetype-based mind map

- Table summary.** It is possible to represent an archetype in detail by using a spreadsheet (Figure 31). Each row represents a class or an attribute of the archetype tree. Different columns are used to inform about the archetype information structure, such as the class type, the occurrences or cardinality, the semantic bindings and data value constraints. The generation of spreadsheets is also generic, independent of the reference model.

Name	Description	Type	Id	Occurrences/Cardinality	Mandatory
Medication use data	Aggregated information about the administration or consumption of a specified medication or type/class of medication at or during an event such as a specific point in time or duration of time.	OBSERVATION	at0000	{1..1}	Y
Event Series events	@ internal @	HISTORY	at0022	{0..1}	N
Defined event data	This event attribute of the reference model needs to be explicitly defined and constrained as a specific point-in-time or interval of time in a template or at run-time.	EVENT	at0023	{0..*}	N
Medication use information items	@ internal @	ITEM_TREE	at0001	{0..1}	N
Medication name value	Name of medication, class or type of medication.	ELEMENT	at0002	{0..1}	N
		DV_TEXT	at0009	{0..1}	Y
Used in event? value	Is the individual using the medication, class or type of medication at or during the identified event?	ELEMENT	at0003	{0..1}	N
		DV_BOOLEAN	at0006	{0..1}	Y
Description value	Narrative description about the use of the medication, class or type of medication at, or during, the identified event.	ELEMENT	at0007	{0..1}	N
		DV_TEXT	at0011	{0..1}	Y
Specific/subclass medication items	Details about a specific medication or subclass of medication, if a medication type or class is specified in the 'Medication name'.	CLUSTER	at0008	{0..*}	N
Medication name value	Name of medication or subclass of medication.	ELEMENT	at0021	{1..*}	Y
		DV_TEXT	at0013	{0..1}	N
Used in event? value	Is the individual using the medication or subclass of medication at or during the identified event?	ELEMENT	at0024	{0..1}	N
		DV_BOOLEAN	at0010	{0..1}	Y
Description value	Narrative description about the use of the medication or subclass of medication at, or during, the identified event.	ELEMENT	at0014	{0..1}	N
				{0..1}	Y

Figure 31. Archetype-based table summary

- Implementation guide document.** Implementation guides are reference documents usually used by technical implementers of EHR standards. They describe the logical requirements to satisfy from an implementation point of view (Figure 32). For example, they indicate how to build a standard XML document following the constraints defined in an archetype. In essence, an implementation guide includes the same information as any archetype, but represented in a technically aimed format.

Medication name

Template:	Template ELEMENT: <i>Medication name (@archetype_node_id: at0002)</i>
Type:	ELEMENT
Contained by:	Template ITEM_TREE: <i>Medication use information (@archetype_node_id: at0001)</i>
Contains:	

Constraint summary

Entity	XPath
Medication name	/data[@archetype_node_id="at0022"]/events/data[@archetype_node_id="at0001"]/items[@archetype_node_id="at0002"]

Description

Name of medication, class or type of medication.

Template Rules

- 1) **SHALL** contain exactly one attribute [1..1] **value**
 - a) Contains at most one [0..1] *DV_TEXT* Class

Figure 32. Archetype-based implementation guide

Another important part of the review process of archetypes is to validate them technically, i.e. to ensure that they are conformant with the RM and with their parent archetypes, in case of being specialized archetypes. LinkEHR includes a validation mechanism that checks all the archetype nodes and provides a summary of the problems detected, such as attributes existence mismatch, objects occurrences discrepancy, and even if the defined regular expressions or constraints are compatible withers defined by the RM. This validation is implemented through a subsumption checking function of labeled nodes, extensively explained in [69].

d) Archetype repository connection

In order to support the complete modeling and governance processes, archetype editors should interact with archetype governance systems. We will describe existing archetype governance systems in detail in Section 6.1.2. LinkEHR Studio

can access to openEHR archetype repositories, known as Clinical Knowledge Manager (CKM), through their public web services [102]. LinkEHR Studio users can search for existing archetypes by the archetype concept name or its header information. The tool will show a list of found archetypes (Figure 33), that we can download and open in LinkEHR Studio automatically, to start reviewing or editing them. The inverse process, uploading new or modified archetypes, is not implemented yet, although it would be desirable in order to coordinate the life cycle and versioning management of the archetypes.

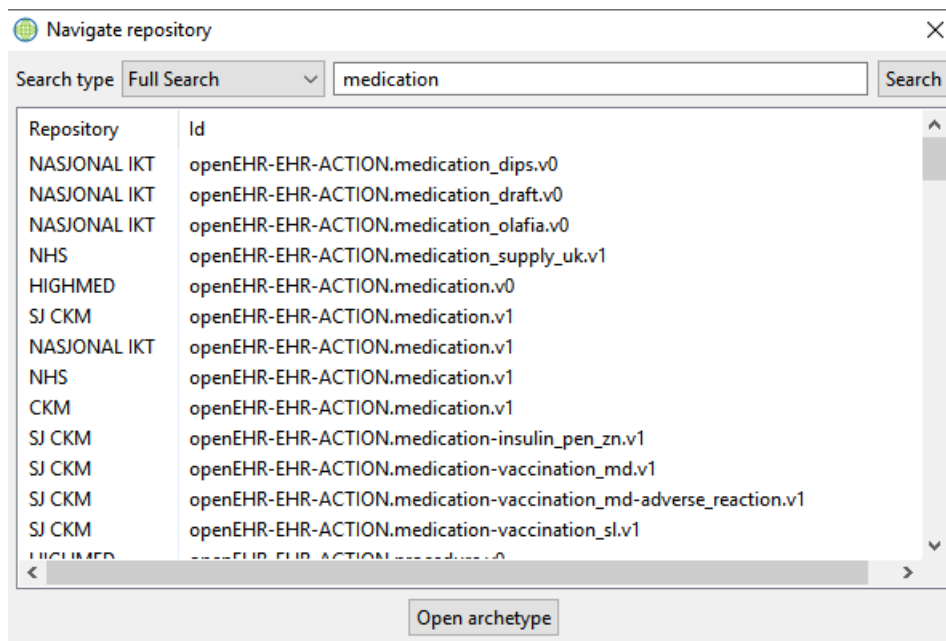


Figure 33. Integration of LinkEHR Studio with archetype repositories

e) Terminology binding and terminology server connection

Technologies for querying and managing clinical terminologies have continuously evolved during the last years. The connection of LinkEHR Studio with terminology servers, in order to complete the archetype definitions, has evolved accordingly. The first implementation in LinkEHR Studio was an API subset of the standard CTS2 web services [28]. In order to ease the implementation, the subset only included those methods and data structures that are useful from an archetype

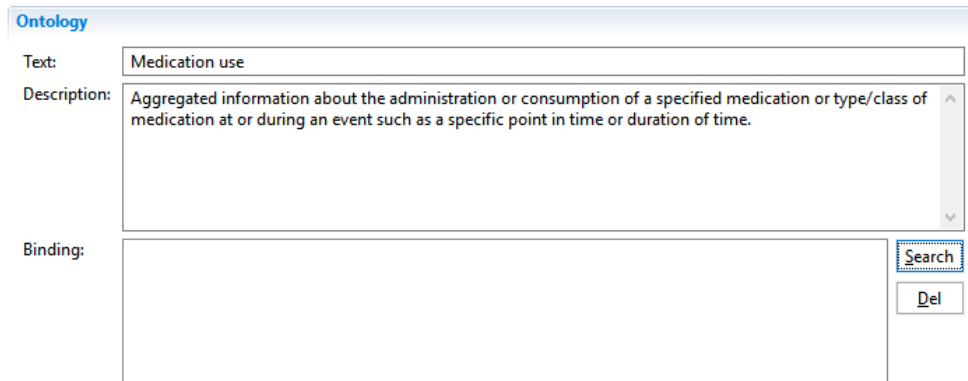
editor perspective. We do not need to support the complete functionality to manage a terminology server. Thus, we can easily access terminology servers that use CTS2. In some cases, we worked with terminology servers that only offered a proprietary API. In those cases, it was easy to implement an adapter from the proprietary API to our CTS2 API subset.

In recent years, HL7 FHIR has become one of the most promising alternatives to communicate clinical data between health information systems using RESTful Web Services. HL7 FHIR includes resources to support the communication of coded concepts and terminology value sets [103]. The ValueSet resource specifies a set of codes drawn from one or more code systems, intended for use in a particular context. This resource allows both the representation of extensional and intensional sets of coded concepts, independently of the terminology used. For example, in Figure 34 we can see an example of two SNOMED CT ValueSets. The first one is an extensional ValueSet, where the codes of the set are explicitly listed. The second example uses the SNOMED CT Expression Constraint Language [26] to define an intensional ValueSet that will be populated on demand. HL7 FHIR ValueSets can be referenced by a unique URL, following the REST approach. Thus, they can be used to define the terminology value mappings in the archetype definition.

Extensional ValueSet	Intensional ValueSet
<pre> { "resourceType": "ValueSet", "id": "example-extensional", ... "compose": { "include": [{ "system": "http://snomed.info/sct", "concept": [{ "code": "266919005", "display": "never smoker" }, { "code": "77176002", "display": "smoker" }, { "code": "8517006", "display": "former smoker" }] }] } } </pre>	<pre> { "resourceType": "ValueSet", "id": "example-intensional", ... "compose": { "include": [{ "system": "http://snomed.info/sct", "filter": [{ "property": "constraint", "op": "=", "value": "< 404684003 Clinical finding : 363698007 Finding site = << 39057004 Pulmonary valve structure " }] }] } } </pre>

Figure 34. Example of HL7 FHIR ValueSets

To define value mappings, but also semantic mappings, LinkEHR Studio includes a specific interface that allows connecting and querying terminology servers. Every class node of the archetype has a binding option in its constraints formulary (Figure 35).



The screenshot shows a window titled "Ontology" with the following fields and controls:

- Text:** A text input field containing "Medication use".
- Description:** A text area containing "Aggregated information about the administration or consumption of a specified medication or type/class of medication at or during an event such as a specific point in time or duration of time." with a vertical scrollbar on the right.
- Binding:** An empty text input field.
- Search:** A button located to the right of the Binding field.
- Del:** A button located below the Search button.

Figure 35. Access to the semantic binding of archetypes

This option gives access to the terminology binding window (Figure 36). Using this window, a clinical user can search for the most appropriate concepts that describe the archetype node (semantic mapping) or bind it to a terminology subset to constrain coded data types (value binding). The user can search directly for a known code, or search for a concept name or description in natural language. The results are shown in a Google instant-search style, being the list of results updated with each keystroke. The user can select one or several codes and map them to the archetype node that is being edited. In case there are not terminology servers available, the user can define a mapping manually, by selecting a terminology and specifying the mapped code.

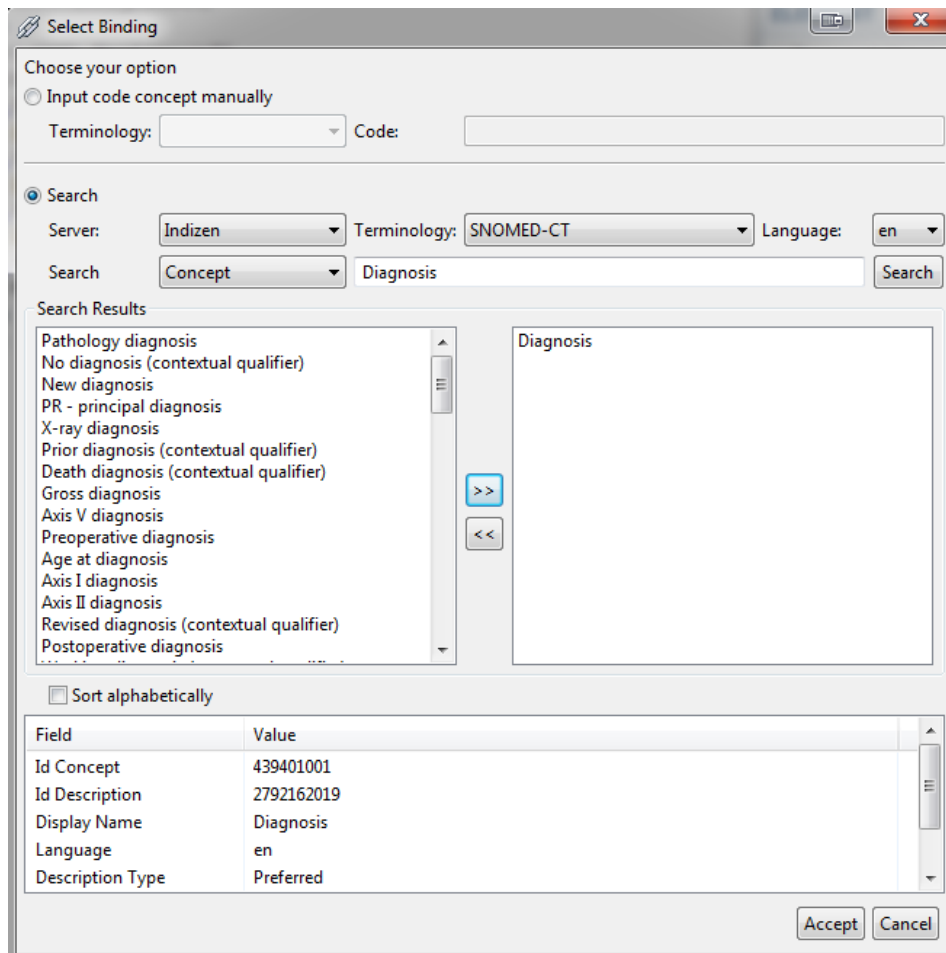


Figure 36. Search windows of terminology concepts and ValueSets

f) Generation of example XML instances and Schematron validation rules

Most EHR standards use XML as the canonical format for data representation. Standard XML documents must be conformant to an XML Schema, which is usually provided as a formal definition format of the standard RM. However, the XML documents must also be conformant to the archetypes that define specific domain constraints.

In order to assist users of LinkEHR Studio with a technical background, the tool provides several functionalities that reduce the gap between the clinical definition of the archetype and the final technical implementations. In particular, LinkEHR Studio provides functionalities to generate examples of XML documents, and to generate validators of those XMLs. The theoretical basis of these developments can be found in D. Boscá PhD Thesis [95].

6.1.2. LinkEHR Model Manager

In addition to the edition of archetypes, we need a platform that supports the governance of all the developed archetypes. Section 3.2.1 describes several CIM and archetype governance systems. In this Section, we will describe in detail LinkEHR Model Manager, which is a result of the archetype governance principles described in this thesis. However, it is important to note that LinkEHR Model Manager is not only an archetype governance system, but a clinical information model governance system, as it will be explained later.

A limitation of most of the governance systems is that they only support one RM and CIM specification format. For example, ART-DÉCOR supports HL7 CDA templates only. Simplifier supports HL7 FHIR profiles only. Finally, all the CKMs support openEHR archetypes only.

Being limited to one RM hinders interoperability, since it complicates the collaboration between experts creating CIM based on different technologies. The relationships between similar concepts represented in different standards are unknown. The possibility of reusing knowledge expressed in different formats is also limited.

We developed LinkEHR Model Manager with the objective of being used for the governance of CIM independently of their RM or specification format. To that end, LinkEHR Model Manager implements the ISO 13972 standard of Detailed Clinical Models (DCM) [104]. This standard, currently under review, “describes requirements for CIMs content, structure and context and specification of their data

elements, data element relationships, meta-data and versioning, and provides guidance and examples". Following those requirements, LinkEHR Model Manager implements an information layer that groups different CIM specification formats under an abstract clinical model definition. It does not imply to have a unique information structure, or to have a transformation of information structures expressed in different standards. It is a solution for governing each of the different RMs and CIM formats under the umbrella of a common clinical model.

LinkEHR is also inspired by ArchMS, an ontology-driven archetype management system [53]. ArchMS is built upon semantic technologies and provides services for archetype search, validation and transformation. However, there are several key differences between ArchMS and LinkEHR Model Manager.

- ArchMS architecture is based on the use of ontologies and semantic representations of archetypes. It provides services that need a prior conversion from archetypes to an OWL representation. In contrast, LinkEHR Model Manager provides a description layer over the stored resources, based on the ISO 13972 standard [104], that provides a common definition format for clinical information models.
- ArchMS was designed with the focus on archetype discovery and transformation, by creating a common ontology over ISO 13606 archetypes and openEHR archetypes. In contrast, LinkEHR Model Manager supports the search of models, the management of their versioning cycle, the management of relationships between models and terminologies, and their publication and integration with other health information systems, such as terminology servers or EHR systems.
- Although ArchMS supports both ISO 13606 and openEHR RM, it only accepts archetypes in ADL as CIM definition format. LinkEHR Model Manager can store and publish any type of representation format, including ADL archetypes, implementation guidelines, documents, terminology reference sets, images, sample clinical documents, etc.

The first version of LinkEHR Model Manager was implemented under the name of a Concept-Oriented Repository (ROC) [105]. It was a module of the Electronic Health Record project of the Valencia regional health service (HSEAVS), in charge of the management of clinical information models. The HSEAVS project used both ISO 13606 and HL7 CDA standards as the basis to build information models and to exchange health data among healthcare information systems. ROC provided support to the management and publication of clinical information models, independently of the specific reference model.

ROC included some of the fundamental ideas of LinkEHR Model Manager, that were further developed later to fulfill the needs of the National EHR project of the Spanish Ministry of Health (HCDSNS). This project defined the information model structures of eight clinical documents shared among all regional health services across Spain. The models were specified using archetypes and ISO 13606, but HL7 CDA was also supported as an alternative implementation. HCDSNS required a governance system for these models that was built upon LinkEHR Model Manager, with the commercial name of Modeling Objects Server (SOM).

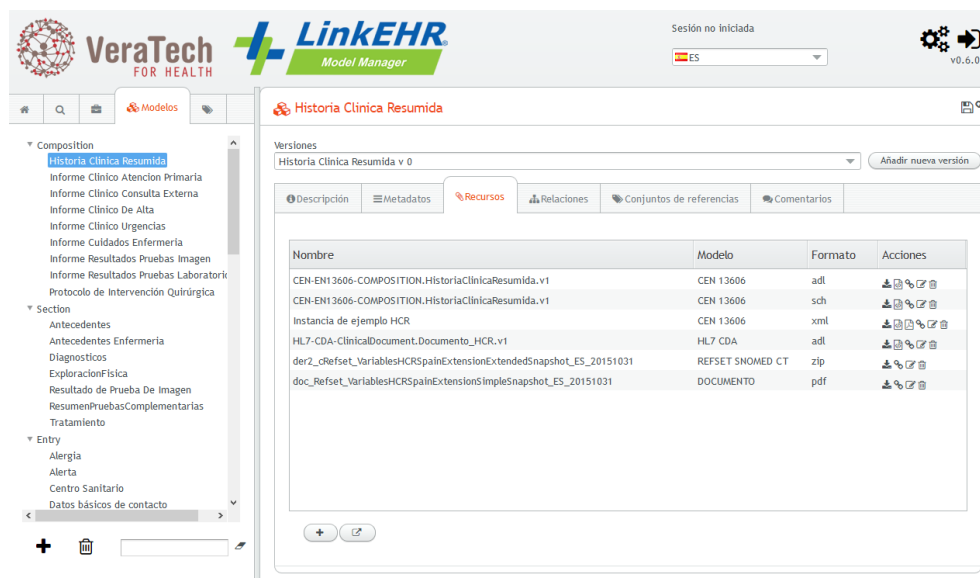


Figure 37. Interface of LinkEHR Model Manager

Taking benefit from the research on archetype governance described in Chapter 5, LinkEHR Model Manager now offers unique characteristics to support the management and publication of CIM, including models specified in archetype form.

The main modules and services of LinkEHR Model manager are represented in Figure 38. The system is centered in the management of clinical information models, which can be versioned. Semantic links can be created between them and any kind of technical resources (such as archetype definitions) can be attached. LinkEHR Model Manager also includes a basic management of terminological reference sets. Finally, projects can be created to group both clinical information models and reference sets. The following section describe all these functionalities in detail.

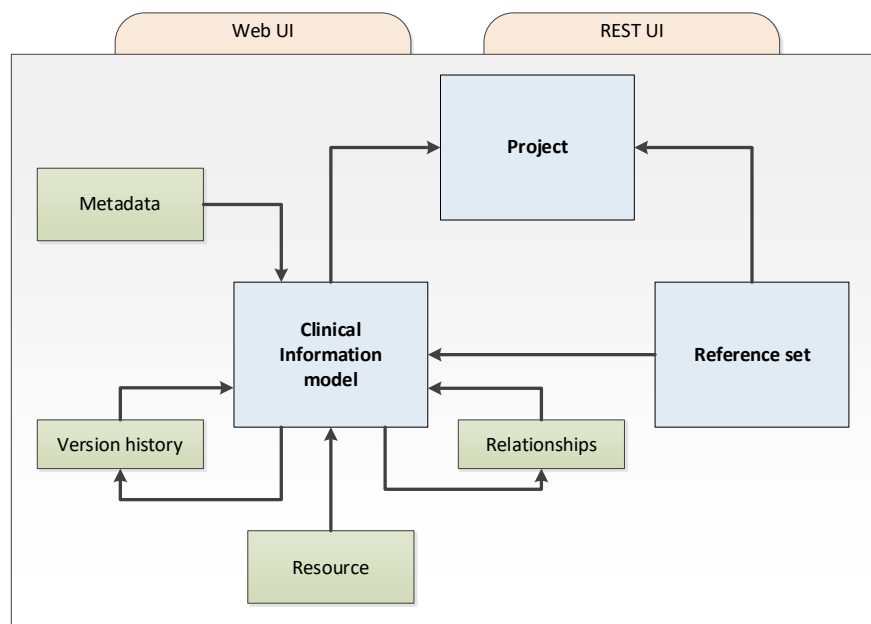


Figure 38. LinkEHR Model Manager modules

a) Clinical information model metadata

The Detailed Clinical Models (DCM) standard defines “the requirements and recommended methods against which clinicians can gather, analyze and specify the

clinical context, content, and structure of Detailed Clinical Models. [...] Describes requirements and principles for DCMs, meta-data, versioning, content and context specification, data element specification and data element relationships” [104].

Following the specifications of the standard, LinkEHR Model Manager allows creating new clinical models accompanied with the multiple metadata. Metadata include fields such as unique name, description in multiple languages, a customizable classification of the model type, version of the model, lifecycle state, authors and contributors, managers, authoring and revision dates, etc. Additionally, it is possible to attach a semantic description of the model, by adding high level correspondences between the model and any terminology or ontology that provide its specific meaning. We can even define multiple correspondences simultaneously. For example, we can describe the clinical model using SNOMED CT concepts or LOINC codes, to make it findable.

b) Resources management

The most important characteristic of LinkEHR Model Manager is its independency of specific standards or CIM specification formats. We can associate any technical or documentation resource to the DCM clinical model. These resources can be specifications in any native format: computable definitions of the CIM such as ADL archetypes, FHIR resources, implementation guidelines, documentation in text or PDF documents, spreadsheets, sample instance files, executable programs that demonstrate the possible use of the model (such as an HTML mockup), XSLT, XQuery or Schematron scripts, etc. All of these resources can be reused or attached to multiple clinical models and can be published by sharing a specific access link to them.

c) Version and lifecycle management

LinkEHR Model Manager allows managing the lifecycle of clinical models from their creation to their revocation. In particular, models can be in one of the following stages: Draft, Team review, Public, Validated, Obsolete, and Revoked. 'Draft' status is used for the first upload or creation of a model which validity is still

unknown. 'Team review' status is used for a model which is being iteratively revised by a team to reach a consensus on it. 'Public' status is used in models that are mature enough to be used but have not been validated in a real system. The latter case is marked with 'Validated' status. Then a model could be marked as 'Obsolete', marking that its use is not recommended anymore and that probably there exist a new version. Finally, 'Revoked' implies that the model has been retired and kept only for historical reference.

Additionally, LinkEHR Model Manager maintains a version history that makes possible to retrieve current or past clinical model definitions and their associated resources. It is important to keep all previous versions of the model and its associated resources to give support to systems that have not yet been updated to the newer specifications.

d) Relationships management

LinkEHR Model Manager provides a mechanism to define relationships between clinical models. A relationship can have a semantic type attached to it, i.e. a code selected from any terminology or ontology. For example, semantic types could be selected from the 246061005 |Attribute (attribute)| hierarchy of SNOMED CT.

Relationships allow to create a semantic network of models, that will facilitate finding related models, or to know the possibly affected models when one of them is updated.

LinkEHR Model Manager defines three types of relationships by default: inclusion, exclusion, and associations. Inclusion indicates that a model is part of the definition of (or that it is contained by) another one. Exclusion indicates the prohibition of containment (one model is not part of the other). Association is a generic relationship that indicates that two models are somehow related. Figure 39 shows a simulated example of model relationships supported by LinkEHR Model Manager.

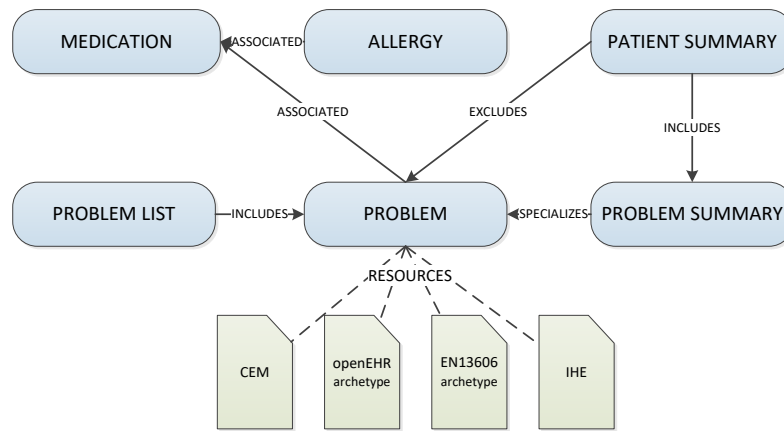


Figure 39. Example relationships between clinical models

In this example, a Problem clinical information model is defined, and several implementations resources are assigned to it, such as the openEHR-EHR-EVALUATION.problem.v1 and CEN-EN13606-ENTRY.problem.v1 archetypes, or the IHE profile of Concern Entry. A medication model is associated to it, and we could define another more precise relationship if we wanted to. Patient summary only contains a problem summary, but not the full problem. Thus, Problem is excluded and Problem summary is included in Patient summary. Problem is included into Problem list. Finally, an additional relationship “Specializes” is defined for Problem summary. This example shows how the relationships can be adapted to the governance needs.

All these kinds of relationships between concepts can be defined through the LinkEHR Model Manager interface and can be also rendered as mind maps, which makes them easier to understand and maintain (Figure 40).

Show relations mindmap

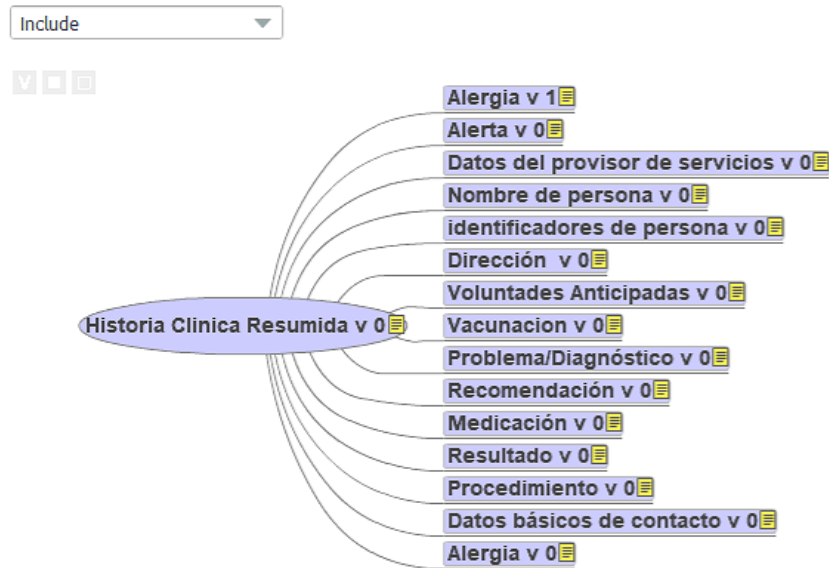


Figure 40. Visualization of model relationships in a mind map

e) Reference set management

As we already know, to complete the semantic definition of clinical information models, we need to associate them to terminology codes and to terminology reference sets. These terminologies should be accessible through a terminology server, where they are controlled and governed. The separation of clinical information model governance and terminology governance has multiple benefits, since they involve different domains of knowledge, tools and professional skills. However, at some point clinical model implementers need to have a complete view of the technical resource, for example the archetype ADL and the reference set member concepts.

LinkEHR Model Manager includes a basic management and publication of terminology reference sets. The objective is to provide a single access point to all the resources needed to implement a clinical information model. In the Reference set section of LinkEHR Model Manager, we can create new reference sets. We can

provide a name and a description to it and select to which terminology it belongs.

Finally, the reference set contents can be defined in four ways:

- **As a URL reference.** This would be the preferred use case when the reference set is available at the terminology server. We only have to provide a link to the location of the actual reference set, to access, review or download it.
- **As an ad hoc subset.** Sometimes, a reference set is a very small collection of concepts. Or it may happen that we don't have access to a terminology server. In this case, we can create an ad hoc reference set directly in LinkEHR Model Manager. It will be a simple list of codes of a terminology. An example could be the creation of a reference set for the gender of patients.
- **As a SNOMED CT expression constraint language.** SNOMED CT provides a specific language called expression constraint language that allows defining intensional reference sets, i.e. the contents of the reference set are defined dynamically according to some properties. We can create an ad hoc reference set using this SNOMED CT language. LinkEHR Model Manager is connected to SNQuery [106], a system that allows the execution of those expressions. As a result, the set of codes compliant with the expression will be retrieved as part of the reference set.
- **As a resource uploaded to the platform.** A last option is to upload a reference set to LinkEHR Model Manager previously created elsewhere. As for a model, we can attach any kind of resource to the reference set containing its concepts: as a spreadsheet file, a CSV, a PDF, a SNOMED CT RF2 file, etc.

Figure 41 depicts the interface for reference set definition in LinkEHR Model Manager.

The screenshot shows a form for defining a reference set. The form has the following fields:

- Name:** SNOMED CT. COVID-19
- Description:** (Empty text area)
- Terminology:** SNOMED (Dropdown menu)
- Substrate:** (Empty dropdown menu)
- Type:** Reference (URI/URL) (Dropdown menu)
- Value:** https://www.mscbs.gob.es/profesionales/hcdsns/areaRecursosSem/snomed-ct/SNOMED_CT_COVID-19.htm (Text input with a copy icon)

Figure 41. Reference set definition in LinkEHR Model Manager

f) Project management

An interoperability scenario, such as the communication of a patient summary for continuity of care or sending information from the EHR to a research information system, will usually require the use of multiple clinical information models and different reference sets. In order to group all the resources to be used in an interoperability use case, LinkEHR Model Manager provides the possibility of creating projects. A project is a container of models and reference sets put together for a specific use. We can associate a name and a description to it, and it can be published as a whole to ease the implementation of the interoperability solution.

g) User and role management

LinkEHR Model Manager includes a comprehensive management of users and roles. It is possible to define new roles in the system and adjust their rights and limitations at a very low level. Figure 42 (in Spanish) shows how we can set different permissions for a specific role, including detailed permissions such as allowing the creation, edition, or removal of models, allowing their versioning, allowing the management of users, or allowing the access to audit information.

Edit role
✕

Name

Comment

Type: System role

Permissions:

Permission	Comment	Set
BORRAR	Permite borrar modelos o versiones	<input type="checkbox"/>
EDITAR	Permite editar el modelo	<input checked="" type="checkbox"/>
DESCARGAR MODELO	Permite llevar a cabo descargas de recursos	<input checked="" type="checkbox"/>
CANCELAR USUARIO	Permite cancelar usuarios	<input type="checkbox"/>
VISTA COMPLETA	Permite ver la vista extendida del modelo	<input checked="" type="checkbox"/>
SUSCRIBIRSE	Permite al usuario suscribirse a modelos	<input checked="" type="checkbox"/>
ESPECIALIZAR	Permite especializar el modelo	<input checked="" type="checkbox"/>
EDITAR INFORMACIÓN DE LOS USUARIOS	Permite editar la información de los usuarios	<input checked="" type="checkbox"/>
AUDITAR	Permite auditar modelos y usuarios	<input type="checkbox"/>
BÚSQUEDA AVANZADA	Permite llevar a cabo la búsqueda avanzada	<input checked="" type="checkbox"/>
INVITAR EDITOR	Invitar editores al modelo	<input checked="" type="checkbox"/>
VERSIONAR	Permite versionar el modelo	<input checked="" type="checkbox"/>
BORRAR USUARIO	Permite borrar usuarios	<input type="checkbox"/>
CREAR	Permite crear modelos o versiones	<input checked="" type="checkbox"/>

Figure 42. Edition of role permissions

Roles are assigned to users of the system, and one user can have multiple roles at the same time. This management allows defining different working groups in the system, such as model creators, reviewers, administrative staff, etc. The functionality offered by the system will vary depending on the role of the logged user.

h) Collaborative governance

LinkEHR Model Manager facilitates the collaboration in the governance of clinical information models.

Any user can subscribe to any model to be notified when any change is produced in it. These notifications will be available in the landing page of the platform just after a user login. There, the user will find a list with the latest events that affect to the subscribed model: changes in its metadata, changes in the associated resources, and changes in their versioning history.

It is also possible to manage rights over a model by using specific roles. The system administrator can create a new role with its specific permissions and assign it to a particular model. Every user with that role will be able to interact with the

model according to the allowed permissions. This approach facilitates the management of groups of users, for example for creating an archetype edition team.

Finally, every model has a comment section. Users can discuss the contents and evolution of the model, in a forum-like interface.

6.2. Use cases

The development and evaluation of the Archetype Modeling Methodology and the companion tools (LinkEHR Studio and LinkEHR Model Manager) is supported by practical experiences that span 10 years of research and development projects. We used initial versions of AMM and the tools in research and commercial projects, where we could apply the methodology, test its practical value, and improve the tools functionalities. Each of these projects provided insights on needed modifications and refinements of the methodology, which were tested in successive projects. The final version of AMM presented in this thesis, together with the latest versions of LinkEHR Studio and LinkEHR Model Manager, are a result of all these experiences. In this section, we summarize several projects that contributed in formalizing the methodology and improving the archetype definition and governance tools.

The first version of AMM was established in 2011 in the regional EHR project of Valencia, Spain. The project objective was to develop a multi-purpose, archetype-based, clinical research database, aggregating clinical data from five million people. We did a requirement analysis and developed the specifications of the contents of the research database in the form of ISO 13606 archetypes and HL7 CDA templates. During the analysis phase, only selected clinical experts from the regional health service participated. This resulted in the definition of very specific and hardly reusable archetypes and showed the need of incorporating a multidisciplinary team of experts. We also learned about the need of developing a conceptual model first, instead of working directly with archetypes implemented in any of the two standards. Otherwise, the continuous changes in the archetypes would have made the project unsustainable. This also allowed avoiding the limitations imposed by EHR

standards in design time. Finally, we learned about the importance of defining the information structures before working with the terminological aspects of the models to ease the development process. Together with the archetype modeling methodology, a first version of LinkEHR Model Manager was developed to help in the management of the archetypes and other semantic resources used in the project.

A second use of the AMM was in the Spanish National EHR project (HCDSNS), started in 2013. This project has the aim of developing an infrastructure that allows sharing relevant health data sets about patients between the 17 regions of Spain. HCDSNS defines ten basic clinical documents: Patient summary, Primary healthcare report, Emergency reports, Discharge reports, External consultation note, Nursing care reports, Laboratory test results, Imaging test results, Medication summary, and Prescriptions. This project demonstrated the importance of having a multidisciplinary group of professionals to define the contents of the documents. In this project 46 professionals from 27 scientific societies, and 21 institutional representatives, participated in the analysis phase. The result was published as a National Rule in the Government Official Journal, specified in free text [107]. Based on that informal specifications of the information models, we designed and developed 64 ISO 13606 archetypes and 36 HL7 CDA archetypes. We did not reuse existing archetypes because no public ISO 13606 or HL7 CDA archetypes existed at that time. However, we analyzed openEHR archetypes from the openEHR CKM as an initial reference. The first release of the archetypes only covered the structural definition of the documents, i.e. they define only the information structures according to the selected RM. In a second phase, the Ministry of Health defined a SNOMED CT national extension to support all the new concepts used in the archetypes, and the appropriate reference sets for data entry points (approximately 70 reference sets). All these codes will be bound to the archetypes. In this project, we also improved LinkEHR Model Manager to better control the lifecycle of archetypes, to allow the definition of relationships between different models, and to include links between models and terminology value sets.

A third relevant project was the construction of a maternal-child care data repository [108]. The aim of this project was to develop an integrated, standardized, and quality assessed, data repository for research and monitoring of best practices of mothers and newborns care. In this project, we followed all phases of the AMM related to the definition of ISO 13606 information structures. Once again, we did not cover the terminological aspects of the archetype definition during the first phase of the project due to the tight schedule and limited budget of the project. The clinical participants of this project had basic to medium knowledge about the ISO 13606 standard and the archetype methodology. This allowed us to develop the analysis and design phases faster than in our previous experiences. The main lesson learned from this project was about the benefits of having members of the work group trained in clinical information modeling. It is not necessary that they have a deep knowledge about EHR standards, but it eases the understanding of the modelling process, and the importance of building generic and reusable archetypes, independently of modeling local templates.

Between 2015 and 2019, the author participated in the National Unified Electronic Health Record project (HCEN) for the National Health System of Uruguay. The aim of the project was to build archetypes and implementation specifications of several clinical documents for the National Unified EHR system, including Patient summary, Primary care outpatient note, Hospital emergency service note, Non-centralized emergency service note, Discharge summary, and Dentistry note. We applied all phases of AMM, including the specification of a governance policy and the deployment of a management system for the resulting archetypes.

The work group was composed by all the suggested profiles of AMM: a clinical group leader, a clinical expert group, a multidisciplinary clinical support team from the National Health System, a terminology expert in SNOMED CT, and a group of technical experts in HL7 CDA, and ISO 13606. All the members participated during different phases of the development and covered all needs of the project. An additional group of final users of the developed specifications also participated to

evaluate the results. This kind of final users was a valuable addition to the work group.

In Phase 1 of the AMM (requirements), clinical experts studied existing documentation related to the use case. They included specifications from existing projects such as the Spanish National EHR specifications, the European epSOS specifications, openEHR archetypes, and the HL7 CDA Implementation Guide for IHE Health Story Consolidation.

In Phase 2 (design), clinical experts designed the information models corresponding to the six clinical documents. To facilitate this task, a template spreadsheet was designed to document the information elements. The spreadsheet allowed including the set of information elements, their naming, their structure, their data types (numbers, texts, dates, coded values...), their cardinality, and the applicable data constraints or code subsets.

In Phase 3 (development), technical experts converted the information model specifications into formal and standardized archetypes. They defined 41 ISO 13606 archetypes, 6 HL7 CDA implementation guides, and sample HL7 CDA data instances. The terminology expert defined SNOMED CT subsets to populate the possible values of coded information elements such as procedures, allergies and diagnosis. In addition, the terminology expert, with the support of the clinical experts, selected the appropriate terminological codes to map and describe the semantics of the information structure of the archetypes. A lesson learned was that it is important to maintain a continuous collaboration between technical and terminology experts during the whole development process. Such a collaboration minimizes the inconsistencies while developing archetypes that are bound to terminologies.

In Phase 4 (validation), we designed sample data entry forms to simulate the data registry process in real clinical environments. Six health institutions, 60 health professionals and more than 120 patients participated in the evaluation process. The overall evaluation of the designed documents was positive, although they also provided suggestions and possible modifications for future revision of the archetypes.

In phase 5 (publication), all the generated materials and specifications, including the information model specifications, the ISO 13606 archetypes, and the HL7 CDA implementation guides were published in LinkEHR Model Manager.

The project scope was limited to the definition of clinical and technical specifications of the six mentioned clinical documents. It will be responsibility of local health information systems implementers to include the new specifications into their future software developments.

Chapter 7

Conclusions and future work

Semantic interoperability of health information is an unquestionable requirement for the sustainability of healthcare, and it is essential to face the new health challenges of a globalized world. This thesis provides new methodologies to tackle some of the fundamental problems of semantic interoperability, specifically those aspects related to the definition and governance of archetypes.

The contributions of the thesis are:

- A **study of existing knowledge modeling methodologies**, focused on software, clinical information model and ontology modeling. This study concluded that there does not exist a consensus about a formal archetype modeling methodology. The study also provided an insight on the activities and good practices that should be part of an archetype modeling methodology. This study contributes to the achievement of objective one (O1) of the thesis.
- A **proposal of a unified Archetype Modeling Methodology (AMM)** based on the previous study. AMM formalizes the phases, the required participants and tools, and the workflow of archetype development. The use of this methodology improves the quality of archetypes, by allowing a consistent definition of multiple archetypes made by different people, all under the same procedures. AMM methodology describes the archetype development process, but it can also serve as a reference for the development of any other type of CIM, for example HL7 CDA templates or HL7 FHIR resources. Different technologies and tools may be used in each case, but the principles of analysis, design development, validation, and publication will remain the same. Archetype designers and developers can adapt the proposed methodology to meet some specific needs of

a particular development project. This contribution fulfils objective two (O2) of the thesis.

- An **analysis of clinical information model governance systems**. This analysis focused on the functionalities provided by existing systems and ongoing initiatives, and their governance strategies, including the study of the supported standards and CIM formats, the management of participants, model metadata, model versioning, model lifecycle, terminologies, technical resources, and clinical evidence documentation. This analysis contributes to the achievement of objective one (O1) of the thesis.
- A **formalization of archetype governance principles** based on the previous analysis. Archetype governance is the framework that guides archetype development. The proposed principles help to define the development strategies, priorities, responsibilities, and guidelines of archetype modeling. A management board oversees the cooperation and coordination of the participants in order to develop interoperable and reusable archetypes. This contribution fulfils objective three (O3) of the thesis.
- **Design and development of tools** that provide support to archetype modeling and archetype governance. These tools, namely LinkEHR Studio and LinkEHR Model Manager, incorporate guidance on all the phases of archetype modeling, and provide a platform for the governance and publication of archetypes, together with other semantic interoperability artifacts, such as terminology value sets and implementation resources. This contribution fulfils objective four (O4) of the thesis.

Multiple projects and development experiences served to put in practice and to evaluate and improve the contributions of this thesis. These experiences vary from local projects inside a single organization that required an accurate reuse of EHR data, to the development of national EHR projects in Spain and Uruguay based on semantic interoperability principles.

There is an international consensus on which are those semantic interoperability principles: the use of EHR reference models, the use of archetypes or any

other type of CIM, and the use of terminologies. This thesis focuses on the development and management of archetypes, but despite its contributions, the following challenges in both archetype modeling and archetype governance still require further research.

- Archetype development and terminology development are closely related. It is necessary to explore the bidirectional influences between both in order to improve the modeling process. The literature describes the boundary problem between archetypes and terminologies [109]. There is a need to achieve a consensus about the limits of each artifact, but it still has not been completely resolved. This is especially relevant in order to represent constraints over fuzzy information [110]. For example, when the representation of age uses a classification instead a numeric value: infant, child, adolescent, adult or elderly. The traditional approach is to assign ranges to each group (an infant is less than 1 year, a child is between 1 and 10, an adolescent is between 10 and 19, etc.). However, these ranges may vary depending on the country, the law, or the reference organization (WHO, UN, National Statistics Office...). Which would be the standard and interoperable definition of this type of information? How can it be consistently used in public health studies or clinical decision support systems?
- Archetypes and terminologies evolve at a different pace. We need to coordinate the independent evolution of the archetype definitions and the terminologies and terminological value sets. Changes on the terminology side can require modifications in the archetype, or vice versa. Can we detect those changes and automatically modify the archetype definition or the value sets? Should we extract terminology bindings, currently embedded in the archetype definition, and manage them using an external mediator? This solution opens new research questions. Can a mediator between archetypes and terminologies adjust dynamically the applicable terminology value set according to the context of use of the archetype?
- During the last years, the conversion and reuse of CIM definitions between different standards and formats is receiving the interest of the research community. There are already experiences and initiatives to build generic and reusable

CIMs [19,104], and research about clinical information models transformation [95]. However, neither solution is preeminent. Health information standards still follow their own formats to define clinical information models, and the reuse and interoperability of those models is still limited. Can we achieve a complete and confident automatic transformation of CIMs? How can we avoid overlapping between CIMs created by different standards or organizations? Is it possible to create EHR information systems that are able to use CIMs from different standards without needing a prior transformation?

- Quality metrics of the resulting archetypes. One of the main benefits of formalizing archetype development and establishing a governance process is to improve the quality of the archetypes. There is already a proposal of quality framework for semantic interoperability definition and implementation [81]. In this work, Moreno-Conde establishes metrics for evaluating semantic interoperability assets quality, which includes archetypes. However, he also concludes that there is a lack of development of governance systems, which have a direct impact on the quality of archetypes. There is a need to continue the research on the implications of better governance methodologies in the quality of the governed archetypes.
- Semantic interoperability is not an end, but a means to create interconnected systems that are able to share and reuse information seamlessly. In this sense, archetype governance should not be limited to the management of archetypes as standalone definitions of clinical information models, but it should also provide support to their use in semantic interoperability implementations. It is common that these implementations go beyond a simple point-to-point communication of data. Instead, clinical data transformation and reasoning workflows are defined in order to deal with different representations and different models of information, and to incorporate more complex processing over existing data [111]. In this context, new research questions arise. How can archetype governance be incorporated to those workflows? How can those workflows locate the archetypes they need? How can they inform about new requirements to the archetype governance?

After millennia of medicine practice, the formalization of the representation of health information has finally received the interest of researchers and developers during the last decades. The aim is to achieve a better representation and reuse of health information, particularly the one existing in EHR information systems. Semantic interoperability is part of the strategy to achieve this objective, and the modeling and governance of archetypes are a means to fulfill it. The contributions of this thesis provide new methodologies and tools to advance in this objective.

Bibliography

- [1] V.N. Stroetmann, D. Kalra, P. Lewalle, A. Rector, J.M. Rodrigues, K.A. Stroetmann, G. Surjan, B. Ustun, M. Virtanen, P.E. Zanstra, Semantic interoperability for better health and safer healthcare, European Commission, Directorate-General Information Society and Media, 2009. <http://dx.doi.org/10.2759/38514>.
- [2] J.L. Whitten, L.D. Bentley, Systems analysis and design methods, 7th ed, McGraw-Hill/Irwin, Boston, 2007.
- [3] ISO/TR 20514:2005 - Health informatics - Electronic health record - Definition, scope and context, (2005). www.iso.org.
- [4] ISO 13606:2008 - Health informatics - Electronic health record communication, (2008).
- [5] ISO 18308:2011 - Health informatics - Requirements for an electronic health record architecture, (2011). www.iso.org.
- [6] HIMSS Europe, Annual European eHealth Survey, (2019). <https://europe.himssanalytics.org/europe/ehealth-barometer/ehealth-trend-barometer-annual-european-ehealth-survey-2019> (accessed June 10, 2020).
- [7] Office of the National Coordinator for Health Information Technology, Office-based Physician Electronic Health Record Adoption, (2019). dash-board.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php (accessed June 10, 2020).
- [8] F.F. Odekunle, S. Srinivasan, R.O. Odekunle, Why Sub-Saharan Africa Lags in Electronic Health Record (EHR) Adoption and Possible Strategies to Increase EHR Adoption in this Region, *Journal of Health Informatics in Africa*. 5 (2018) 8–15. <https://doi.org/10.12856/JHIA-2018-v5-i1-147>.
- [9] Healthcare Information and Management Systems Society, What is Interoperability?, (2019). <https://www.himss.org/library/interoperability-standards/what-is-interoperability> (accessed June 10, 2020).

- [10] European Commission, eHealth Digital Service Infrastructure, (2020). <https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home> (accessed June 10, 2020).
- [11] A. Moreno-Conde, D. Moner, W.D. da Cruz, M.R. Santos, J.A. Maldonado, M. Robles, D. Kalra, Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis, *Journal of the American Medical Informatics Association*. (2015). <https://doi.org/10.1093/jamia/ocv008>.
- [12] openEHR, (n.d.). <http://www.openehr.org/> (accessed July 18, 2014).
- [13] ISO 13606:2019 - Health informatics - Electronic health record communication, (2019). www.iso.org.
- [14] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Behlen, P.V. Biron, A.S. Shvo, HL7 Clinical Document Architecture, Release 2, *J Am Med Inform Assoc*. 13 (2006) 30–39. <https://doi.org/10.1197/jamia.M1888>.
- [15] HL7 International, Fast Healthcare Interoperability Resources (FHIR), (n.d.). <http://hl7.org/fhir/> (accessed August 27, 2020).
- [16] W. Goossen, A. Goossen-Baremans, M. van der Zel, Detailed Clinical Models: A Review, *Healthcare Informatics Research*. 16 (2010) 201. <https://doi.org/10.4258/hir.2010.16.4.201>.
- [17] G.W. Beeler, HL7 Version 3—An object-oriented methodology for collaborative standards development, *International Journal of Medical Informatics*. 48 (1998) 151–161. [https://doi.org/10.1016/S1386-5056\(97\)00121-4](https://doi.org/10.1016/S1386-5056(97)00121-4).
- [18] T. Beale, Archetypes: Constraint-based Domain Models for Futureproof Information Systems, *OOPSLA 2002, Workshop on Behavioural Semantics*. (2002).
- [19] Clinical Information Modeling Initiative (CIMI), (n.d.). <http://www.hl7.org/Special/Committees/cimi/index.cfm> (accessed March 27, 2017).
- [20] W.T. Goossen, Using detailed clinical models to bridge the gap between clinicians and HIT, *Stud Health Technol Inform*. 141 (2008) 3–10.

-
- [21] T.A. Oniki, J.F. Coyle, C.G. Parker, S.M. Huff, Lessons learned in detailed clinical modeling at Intermountain Healthcare, *J Am Med Inform Assoc.* (2014) amiajnl-2014-002875. <https://doi.org/10.1136/amiajnl-2014-002875>.
- [22] A.L. Rector, Clinical terminology: why is it so hard?, (n.d.). <https://www.ncbi.nlm.nih.gov/pubmed/10805008> (accessed April 8, 2019).
- [23] SNOMED International, SNOMED CT - Systematized Nomenclature of Medicine - Clinical Terms, (n.d.). <http://www.snomed.org> (accessed July 13, 2020).
- [24] WHO | International Classification of Diseases (ICD) Information Sheet, WHO. (n.d.). <https://www.who.int/classifications/icd/factsheet/en/> (accessed July 10, 2019).
- [25] LOINC — The freely available standard for identifying health measurements, observations, and documents., (n.d.). <https://loinc.org/> (accessed July 10, 2019).
- [26] Expression Constraint Language - Specification and Guide, (2017). <https://confluence.ihtsdotools.org/display/DOCECL/Expression%2BConstraint%2BLanguage%2B-%2BSpecification%2Band%2BGuide> (accessed October 2, 2018).
- [27] A.L. Rector, W.D. Solomon, W.A. Nowlan, T.W. Rush, P.E. Zanstra, W.M. Claassen, A Terminology Server for medical language and medical information systems, *Methods Inf Med.* 34 (1995) 147–157.
- [28] Common Terminology Services 2 (CTS2), v1.2, (2015). <http://www.omg.org/spec/CTS2> (accessed July 31, 2019).
- [29] W. Royce, Managing the Development of Large Software System, *Proceedings of IEEE WESCON 26.* (1970) 1–9.
- [30] C. Larman, V.R. Basili, Iterative and incremental developments. a brief history, *Computer.* 36 (2003) 47–56. <https://doi.org/10.1109/MC.2003.1204375>.
- [31] B. Boehm, A Spiral Model of Software Development and Enhancement, *SIG-SOFT Softw. Eng. Notes.* 11 (1986) 14–24. <https://doi.org/10.1145/12944.12948>.

- [32] P. Abrahamsson, J. Warsta, M.T. Siponen, J. Ronkainen, New directions on agile methods: a comparative analysis, in: *IEEE*, 2003: pp. 244–254. <https://doi.org/10.1109/ICSE.2003.1201204>.
- [33] 1074-2006 - IEEE Standard for Developing a Software Project Life Cycle Process, (2006). www.ieee.org.
- [34] ISO/IEC 12207:2008 - Systems and software engineering - Software life cycle processes, (2008). www.iso.org.
- [35] A. Chmura, J.M. Heumann, *Logical Data Modeling - What it is and How to do it*, Springer, 2005. <http://www.springer.com/in/book/9780387229508> (accessed July 20, 2017).
- [36] A. De Nicola, M. Missikoff, R. Navigli, A software engineering approach to ontology building, *Information Systems*. 34 (2009) 258–275. <https://doi.org/10.1016/j.is.2008.07.002>.
- [37] Y. Sure, S. Staab, R. Studer, O. Gmbh, *On-To-Knowledge Methodology (OTKM)*, in: *Handbook on Ontologies*, International Handbooks on Information Systems, Springer, 2003: pp. 117–132.
- [38] M. Grüninger, M.S. Fox, *Methodology for the Design and Evaluation of Ontologies*, in: *Proceedings IJCAI-95*, Montreal, 1995.
- [39] *IDEF5 Method Report*, Armstrong Laboratory, Wright-Patterson Air Force Base, Ohio, United States, 1994. <http://www.idef.com/pdf/Idef5.pdf>.
- [40] M. Fernández-López, A. Gómez-Pérez, N. Juristo, *METHONTOLOGY: From Ontological Art Towards Ontological Engineering*, in: *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*, Facultad de Informática (UPM), Stanford University, EEUU, 1997. <http://oa.upm.es/5484/> (accessed May 8, 2015).
- [41] J. Buck, S. Garde, C.D. Kohl, P. Knaup-Gregori, Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach, *International Journal of Medical Informatics*. 78 (2009) 521–531. <https://doi.org/10.1016/j.ijmedinf.2009.03.001>.

-
- [42] W. Van Grembergen, *Strategies for Information Technology Governance*, IGI Global, 1AD. <https://www.igi-global.com/book/strategies-information-technology-governance/952> (accessed June 12, 2019).
- [43] P.W. and J.W. Ross, *A Matrixed Approach to Designing IT Governance*, MIT Sloan Management Review. 46 (2005AD) 26–34.
- [44] S. Garde, S. Heard, E.J. Hovenga, *Archetypes in Electronic Health Records: Making the Case and Showing the Path for Domain Knowledge Governance*, HIC 2005 and HINZ 2005: Proceedings. (2005) 267.
- [45] S. Garde, *Clinical knowledge governance: the international perspective*, Stud Health Technol Inform. 193 (2013) 269–281.
- [46] W. Antje, H. Birger, M. Michael, *Clinical Knowledge Governance Framework for Nationwide Data Infrastructure Projects*, Studies in Health Technology and Informatics. (2018) 196–203. <https://doi.org/10.3233/978-1-61499-858-7-196>.
- [47] C.S. A, G. Emily, M. Steven, M. Perry, M. Saverio, C. Doreen, T. Diana, R. Roberto, *A Practical Approach to Governance and Optimization of Structured Data Elements*, Studies in Health Technology and Informatics. (2015) 7–11. <https://doi.org/10.3233/978-1-61499-564-7-7>.
- [48] openEHR Foundation, *Clinical Knowledge Manager*, (n.d.). <http://openehr.org/ckm/> (accessed March 27, 2020).
- [49] Nasjonal IKT, *Nasjonal IKT Clinical Knowledge Manager*, (2020). <https://arketyper.no/ckm/> (accessed June 25, 2020).
- [50] HiGHmed, *HiGHmed Clinical Knowledge Manager*, (2020). <https://ckm.highmed.org/ckm/> (accessed June 25, 2020).
- [51] H. Leslie, *State of the clinical modelling program and international CKM*, (2019). <https://www.atomicainformatics.com/archetypical/2019/7/23/state-of-the-clinical-modelling-program-and-international-ckm> (accessed September 14, 2020).
- [52] Healthcare Modeling Collaboration, *Healthcare Modeling Collaboration archetype manager*, (2020). <http://hmc.openehr.org.cn> (accessed June 25, 2020).

- [53] M. del C. Legaz-García, Cc. Martínez-Costa, M. Menarguez-Tortosa, J.T. Fernández-Breis, Exploitation of ontologies for the management of clinical archetypes in ArchMS, in: KR-MED Series CEUR Workshop Proceedings, Graz, Austria, 2012. <http://ceur-ws.org/Vol-897/session3-paper16.pdf>.
- [54] Marand, Archetype designer, (2020). <https://tools.openehr.org/designer> (accessed September 10, 2020).
- [55] M. Kohler, C. Rinner, G. Hübner-Bloder, S. Saboor, E. Ammenwerth, G. Duftschmid, The Archetype-enabled EHR system ZK-ARCHE - integrating the ISO/EN 13606 standard and IHE XDS profile, *Stud Health Technol Inform.* 169 (2011) 799–803.
- [56] R. Lozano-Rubí, A. Muñoz Carrero, P. Serrano Balazote, X. Pastor, OntoCR: A CEN/ISO-13606 clinical repository based on ontologies, *Journal of Biomedical Informatics.* 60 (2016) 224–233. <https://doi.org/10.1016/j.jbi.2016.02.007>.
- [57] ART-DECOR, (2020). <https://art-decor.org/> (accessed June 6, 2020).
- [58] Firely, SIMPLIFIER.NET, (2020). <https://simplifier.net/> (accessed June 26, 2020).
- [59] Clinical Information Modeling Initiative, CIMI Browser, (2020). <http://models.opencimi.org/> (accessed June 26, 2020).
- [60] Clinical Element Model, OpenCEM Browser, (2020). <http://www.opencem.org/> (accessed June 26, 2020).
- [61] Ministerio de Sanidad, Servidor de Objetos de Modelado del Sistema Nacional de Salud (SOM-SNS), (2020). <https://www.somsns.es/> (accessed June 26, 2020).
- [62] National Center for Biomedical Ontology, NCBO BioPortal, (2020). <http://bioportal.bioontology.org/> (accessed June 28, 2020).
- [63] D. Moner, J.A. Maldonado, M. Robles, Archetype modeling methodology, *Journal of Biomedical Informatics.* 79 (2018) 71–81. <https://doi.org/10.1016/j.jbi.2018.02.003>.
- [64] T. Beale, Archetypes: Constraint-based domain models for future-proof information systems, in: Eleventh OOPSLA Workshop on Behavioral Semantics:

-
- Serving the Customer, Northeastern University, Seattle, Washington, USA, 2002: pp. 16–32.
- [65] D. Kalra, A. Tapuria, T. Austin, G. De Moor, Quality requirements for EHR archetypes, in: 2012: pp. 48–52. <https://doi.org/10.3233/978-1-61499-101-4-48>.
- [66] M. Menárguez-Tortosa, J.T. Fernández-Breis, OWL-based reasoning methods for validating archetypes, *Journal of Biomedical Informatics*. 46 (2013) 304–317. <https://doi.org/10.1016/j.jbi.2012.11.009>.
- [67] D. Kalra, A. Tapuria, G. Freriks, F. Mennerat, J. Devlies, Management and maintenance policies for EHR interoperability resources [36 pages], The European Commission, Brussels, 2008. <http://discovery.ucl.ac.uk/66229/> (accessed May 8, 2015).
- [68] A. Moreno-Conde, G. Thienpont, I. Lamote, P. Coorevits, C. Parra, D. Kalra, European Interoperability Assets Register and Quality Framework Implementation, *Studies in Health Technology and Informatics*. (2016) 690–694. <https://doi.org/10.3233/978-1-61499-678-1-690>.
- [69] J.A. Maldonado, D. Moner, D. Boscá, J.T. Fernández-Breis, C. Angulo, M. Robles, LinkEHR-Ed: a multi-reference model archetype editor based on formal semantics, *Int J Med Inform*. 78 (2009) 559–570. <https://doi.org/10.1016/j.ijmedinf.2009.03.006>.
- [70] H. Leslie, Archetype content and style guide, (2020). <https://openehr.atlasian.net/wiki/spaces/healthmod/pages/304742407/Archetype+content+style+guide> (accessed July 15, 2020).
- [71] Ocean Informatics, openEHR Archetype Editor, openEHR Foundation, n.d. <http://openehr.org/downloads/archetypeeditor/home>.
- [72] Archetype Definition Language 2 (ADL2) Specification, (n.d.). <http://www.openehr.org/releases/AM/latest/docs/ADL2/ADL2.html> (accessed March 30, 2017).
- [73] A. Rector, R. Qamar, T. Marley, Binding ontologies and coding systems to electronic health records and messages, *Applied Ontology*. (2009) 51–69. <https://doi.org/10.3233/AO-2009-0063>.

- [74] A. Coenen, T.Y. Kim, Development of terminology subsets using ICNP®, *International Journal of Medical Informatics*. 79 (2010) 530–538. <https://doi.org/10.1016/j.ijmedinf.2010.03.005>.
- [75] R. Qamar, J. (Subbarao) Kola, A.L. Rector, Unambiguous data modeling to ensure higher accuracy term binding to clinical terminologies, *AMIA Annu Symp Proc*. 2007 (2007) 608–613.
- [76] E. Sundvall, R. Qamar, M. Nyström, M. Forss, H. Petersson, D. Karlsson, H. Åhlfeldt, A. Rector, Integration of tools for binding archetypes to SNOMED CT, *BMC Medical Informatics and Decision Making*. 8 (2008) S7. <https://doi.org/10.1186/1472-6947-8-S1-S7>.
- [77] R. Hausam, ed., *HL7 Version 3 Implementation Guide: TermInfo - Using SNOMED CT in CDA R2 Models, Release 1*, (2015). http://www.hl7.org/implement/standards/product_brief.cfm?product_id=418 (accessed September 14, 2020).
- [78] Archetype Identification specification, (n.d.). http://www.openehr.org/releases/AM/latest/docs/Identification/Identification.html#_archetype_identification_specification (accessed March 27, 2017).
- [79] openEHR, openEHR COVID-19 Project, (2020). <https://discourse.openehr.org/t/openehr-covid-19-project/448> (accessed June 12, 2020).
- [80] L.B. Silje, National governance of archetypes in Norway, *Studies in Health Technology and Informatics*. (2015) 1091–1091. <https://doi.org/10.3233/978-1-61499-564-7-1091>.
- [81] A. Moreno-Conde, *Quality framework for semantic interoperability in health informatics: definition and implementation*, University College London, 2016. https://www.researchgate.net/publication/321462340_Quality_framework_for_semantic_interoperability_in_health_informatics_definition_and_implementation (accessed July 7, 2020).
- [82] Project Management Institute, *Project Management Professional Certification (PMP)*, (n.d.). <https://www.pmi.org/certifications/types/project-management-pmp> (accessed July 8, 2020).

-
- [83] Axelos, Projects in controlled environments (PRINCE2), (n.d.). <https://www.axelos.com/best-practice-solutions/prince2/what-is-prince2> (accessed July 8, 2020).
- [84] openEHR - ADL Workbench Home, (n.d.). <https://www2.openehr.org/downloads/ADLworkbench/home> (accessed July 24, 2019).
- [85] T. Beale, Basic Meta-Model (BMM), (2019). <https://specifications.openehr.org/releases/LANG/latest/bmm.html> (accessed July 24, 2019).
- [86] D. Moner, J.A. Maldonado, D. Bosca, J.T. Fernandez, C. Angulo, P. Crespo, P.J. Vivancos, M. Robles, Archetype-Based Semantic Integration and Standardization of Clinical Data, in: Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, IEEE, New York, 2006: pp. 5141–5144. <https://doi.org/10.1109/IEMBS.2006.260124>.
- [87] G. Dufts Schmid, C. Rinner, M. Kohler, G. Huebner-Bloder, S. Saboor, E. Ammenwerth, The EHR-ARCHE project: Satisfying clinical information needs in a Shared Electronic Health Record System based on IHE XDS and Archetypes, *International Journal of Medical Informatics*. 82 (2013) 1195–1207. <https://doi.org/10.1016/j.ijmedinf.2013.08.002>.
- [88] K. Georgy, H. Claudia, D. Hans, Applying UML Connectors for Arranging Medical Archetypes into a Knowledge Base, *Studies in Health Technology and Informatics*. (2013) 1097–1097. <https://doi.org/10.3233/978-1-61499-289-9-1097>.
- [89] F. Klasson, P. Väyrynen, Development of an API for creating and editing openEHR archetypes, (2009). <https://www.diva-portal.org/smash/get/diva2:210297/FULLTEXT01.pdf> (accessed May 28, 2020).
- [90] M. A, M. A, N. F, P. C, Testing a New CDA Development Strategy in a Standardized Infrastructure for Continuity of Care in Chronic Kidney Disease Patients, *Ejbi*. (2012). <https://doi.org/10.24105/ejbi.2012.08.3.6>.

- [91] Z.S.N. Reis, J. d S. Gaspar, A.C. d Souza, M.R. d S. Junior, T.A. Maia, M.R. d Santos, Electronic Systems Interoperability Study: Based on the Interchange of Hospital Obstetrical Information, in: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, 2015: pp. 201–204. <https://doi.org/10.1109/CBMS.2015.57>.
- [92] C. Rinner, M. Kohler, G. Hübner-Bloder, S. Saboor, E. Ammenwerth, G. Duftschmid, Creating ISO/EN 13606 archetypes based on clinical information needs, in: Proceedings of EFMI Special Topic Conference" e-Health Across Borders Without Boundaries, 2011: pp. 14–15. <http://docs.google.com/viewer?url=http://www.meduni-wien.ac.at/msi/mias/papers/Rinner2011a.pdf> (accessed July 15, 2014).
- [93] D.N. Spigolon, C.M.C. Moro, Nursing Minimum Data Set Based on EHR Archetypes Approach, Nurs Inform. 2012 (2012). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3799189/> (accessed June 29, 2015).
- [94] D.N. Spigolon, C.M.C. Moro, Essential data set's archetypes for nursing care of endometriosis patients, Rev Gaucha Enferm. 33 (2012) 22–32.
- [95] D. Boscá Tomás, Detailed clinical models and their relation with electronic health records, Universitat Politècnica de València, 2016. <https://ri-unet.upv.es/handle/10251/62174> (accessed July 26, 2019).
- [96] D. Moner, J.A. Maldonado, D. Boscá, A. Mañas, M. Robles, Development of a Visual Editor for the Definition of HL7 CDA Archetypes, in: L.M. Roa Romero (Ed.), XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, Springer International Publishing, Cham, 2014: pp. 1258–1261. http://link.springer.com/10.1007/978-3-319-00846-2_311 (accessed October 1, 2013).
- [97] D. Bosca, D. Moner, J.A. Maldonado, M. Robles, Combining Archetypes with Fast Health Interoperability Resources in Future-proof Health Information Systems, Stud Health Technol Inform. 210 (2015) 180–184.

-
- [98] ASTM E2369 - 12 Standard Specification for Continuity of Care Record (CCR), (n.d.). <https://www.astm.org/Standards/E2369.htm> (accessed July 29, 2019).
- [99] HL7/ASTM Implementation Guide for CDA® R2 - Continuity of Care Document (CCD®) Release 1, (n.d.). https://www.hl7.org/implement/standards/product_brief.cfm?product_id=6 (accessed July 29, 2019).
- [100] ODM-XML, (n.d.). <https://www.cdsc.org/standards/data-exchange/odm> (accessed July 29, 2019).
- [101] MedXML Consortium, (n.d.). <http://www.medxml.net/> (accessed July 29, 2019).
- [102] CKM REST API, (n.d.). <https://ckm.openehr.org/ckm/rest-doc/#/> (accessed July 30, 2019).
- [103] Terminology-module - FHIR v4.0.0, (n.d.). <http://hl7.org/fhir/terminology-module.html> (accessed July 31, 2019).
- [104] ISO/TS 13972:2015 - Health informatics - Detailed clinical models, characteristics and processes, (2015). www.iso.org.
- [105] D. Bosca, L. Marco, D. Moner, J.A. Maldonado, L. Insa, M. Robles, Detailed Clinical Models Governance System in a Regional EHR Project, in: L.M. Roa Romero (Ed.), XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, Springer International Publishing, Cham, 2014: pp. 1266–1269. http://link.springer.com/10.1007/978-3-319-00846-2_313 (accessed October 1, 2013).
- [106] V.M. Giménez-Solano, J.A. Maldonado, S. Salas-García, D. Boscá, M. Robles, Implementation of an Execution Engine for SNOMED CT Expression Constraint Language, *Stud Health Technol Inform.* 228 (2016) 466–470.
- [107] Ministerio de Sanidad y Política Social, Real Decreto 1093/2010, de 3 de septiembre, por el que se aprueba el conjunto mínimo de datos de los informes clínicos en el Sistema Nacional de Salud, 2010. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2010-14199.
- [108] C. Sáez, D. Moner, R. García-De-León-Chocano, V. Muñoz-Soler, R. García-De-León-González, J.A. Maldonado, D. Boscá, S. Tortajada, M. Robles, J.M. García-

- Gómez, M. Alcaraz, P. Serrano, J.L. Bernal, J. Rodríguez, G. Bustos, M. Esparza, A Standardized and Data Quality Assessed Maternal-Child Care Integrated Data Repository for Research and Monitoring of Best Practices: A Pilot Project in Spain, *Studies in Health Technology and Informatics*. (2017) 539–543. <https://doi.org/10.3233/978-1-61499-753-5-539>.
- [109]D. Markwell, L. Sato, E. Cheetham, Representing Clinical Information using SNOMED Clinical Terms with Different Structural Information Models., in: *KR-MED*, 2008: pp. 72–79. <http://www.kr-med.org/2008/proceedings/proceedings-kr-med2008.pdf#page=77> (accessed May 7, 2015).
- [110]R. Seising, M.E. Tabacchi, eds., *Fuzziness and Medicine: Philosophical Reflections and Application Systems in Health Care*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. <https://doi.org/10.1007/978-3-642-36527-0>.
- [111]J.A. Maldonado, M. Marcos, J.T. Fernández-Breis, V.M. Giménez-Solano, M. del C. Legaz-García, B. Martínez-Salvador, CLIN-IK-LINKS: A platform for the design and execution of clinical data transformation and reasoning workflows, *Computer Methods and Programs in Biomedicine*. (2020) 105616. <https://doi.org/10.1016/j.cmpb.2020.105616>.