Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis

Alberto Moreno-Conde1,2,+, David Moner3,+, Wellington Dimas da Cruz4, Marcelo R Santos5, José Alberto Maldonado3,6, Montserrat Robles3 and Dipak Kalra1

ABSTRACT

Objective This systematic review aims to identify and compare the existing processes and methodologies that have been published in the literature for defining clinical information models (CIMs) that support the semantic interoperability of electronic health record (EHR) systems.

Material and Methods Following the preferred reporting items for systematic reviews and meta-analyses systematic review methodology, the authors reviewed published papers between 2000 and 2013 that covered that semantic interoperability of EHRs, found by searching the PubMed, IEEE Xplore, and ScienceDirect databases. Additionally, after selection of a final group of articles, an inductive content analysis was done to summarize the steps and methodologies followed in order to build CIMs described in those articles.

Results Three hundred and seventy-eight articles were screened and thirty six were selected for full review. The articles selected for full review were analyzed to extract relevant information for the analysis and characterized according to the steps the authors had followed for clinical information modeling.

Discussion Most of the reviewed papers lack a detailed description of the modeling methodologies used to create CIMs. A representative example is the lack of description related to the definition of terminology bindings and the publication of the generated models. However, this systematic review confirms that most clinical information modeling activities follow very similar steps for the definition of CIMs. Having a robust and shared methodology could improve their correctness, reliability, and quality.

Conclusion Independently of implementation technologies and standards, it is possible to find common patterns in methods for developing CIMs, suggesting the viability of defining a unified good practice methodology to be used by any clinical information modeler.

Keywords: electronic health record, semantic interoperability; clinical information modeling; systematic review; clinical information model

BACKGROUND AND SIGNIFICANCE

The increased adoption of electronic health record (EHR) systems potentially enables sharing patient information across multiple systems to support continuity of care. To this end, standards and technical specifications have been developed, which define how the information contained in EHRs should be structured, semantically described, and communicated. Current trends followed by most of those specifications rely on differentiating the representation of data instances from the definition of clinical information models (CIMs).

Clinical Information Models

In this paper, we use the expression “CIM” as a generic term that encompasses all technical specifications defining how clinical information is organized and described inside an EHR system or repository, or for EHR communication. A CIM defines both the information structure and formal semantics of documented clinical concepts. CIMs are structural and semantic artefacts 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 (RM), which provides the basic characteristics and attributes needed to represent data instances. Terminologies such as SNOMED CT, ICD, or LOINC also play an important role in defining CIMs. The structure of CIMs can be bound (precisely mapped) to clinical terminologies to provide a universal definition of the model. Furthermore, terminologies are also used to specify value sets, i.e., the set of possible terms that can be assigned as clinical information values. Thus, a complete semantically interoperable definition of CIMs can only be achieved by using both a standard RM and terminologies to describe the semantics of
the information structures. Goossen et al. descried initiatives that follow a CIM approach, indicating their differences and similarities.

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

HL7 Fast Health Interoperability Resources is a new generation specification that 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 for health information systems. They are, therefore, also CIMs, to some extent.

Another important modeling approach is based on the dual level methodology, which is, in turn, based on the definition of a synthesized and generic RM designed to represent the most basic properties and structures of any EHR. CIMs are defined in the form of archetypes. Archetypes define how data should be structured in order to be seamlessly stored in or transferred between EHR systems. The dual model approach is supported by the EN ISO 13606 standard and openEHR specifications.

Additional modeling approaches focused on defining generic information models at a conceptual level, without depending on a specific implementation, have emerged. The Clinical Information Modeling Initiative, Detailed Clinical Models, and Clinical Element Model Specification are examples of such generic models.

Figure 1 summarizes the reference models used (ie, the models that represent data instances) and the CIM technology employed by each of these initiatives.

Clinical Information Modeling Processes
We define a clinical information modeling process (CIMP) as the process of analyzing the domain and requirements, as well as designing, implementing, validating, and maintaining CIMs. This process will usually require that technical and clinical experts cooperate with one another, in order to obtain a final implementable definition of CIMs that satisfies clinical needs, a definition that may be agreed upon and used at the level of a single care organization, an EHR system vendor user group, a health region, or a country. Once CIMs are defined, governance mechanisms can be applied to ensure correct management and future evolution of the defined models.

The traditional software development process includes requirements definition, a domain analysis, design, implementation, and validation. CIMPs covers domain analysis, design, implementation, and validation of the CIMs but also includes some special characteristics. CIMs are based on standard specifications and formats, and they can be shared and reused. The participation of both health and technical professionals in this process requires coordination and evaluation mechanisms in order to encourage trust in the developed CIMs. Moreover, having a well-defined CIMP is of extreme importance in order to ensure comparable quality and homogeneous design of CIMs created by different organizations or professionals.

OBJECTIVE
This paper aims to identify and compare existing CIMPs and methodologies that have been published in the literature. In particular, a systematic review and an inductive content analysis have been performed in order to learn about methodologies and experiences in building CIMs for semantically interoperable EHR systems. The question being addressed in this study is whether an emergent consensus (good practice) strategy for building CIM artefacts exists, and, therefore, if it is possible to propose a common or unified CIMP.

MATERIALS AND METHODS

Systematic Review
In order to perform a systematic review of the existing literature, we chose preferred reporting items for systematic reviews and meta-analyses methodology. This methodology proposes a 27-item checklist and a flow chart to guide authors during the conduct of a systematic review.

The eligibility criteria (ie, the characteristics to be taken into account to perform the search) were:

- Papers with any of the following terms in their title or abstract: “Electronic Health Record,” “Hospital Information System,”
available in the titles and abstracts, the papers were accepted for full review. In Phase 2 (full review), we reviewed the full text of the selected papers. The objective of this full review was two-fold: to reject those papers that did not fit the purpose of the systematic review and, from only those papers that were finally accepted, to extract a set of data items and indicators to perform further analysis.

Inductive Analysis
In addition to the systematic review, we applied a methodology called inductive content analysis to extract the CIMP steps described in the selected papers. Using this methodology is recommended to avoid creating preconceived categories when the existing literature is limited or heterogeneous. According to this methodology, a set of tags that qualify the CIM definition processes described in the papers were iteratively refined to represent an abstraction of CIMP steps. Information about the modeling processes was organized into categories, in order to provide a high-level and summarized description of those steps.

RESULTS
As a result of the search, 374 papers were found, 18 of which were duplicates. Additionally, the authors identified four additional references that met the search criteria and were relevant to the review, but that were not indexed by the search engines. In total, 369 paper titles and abstracts were screened by the authors, and 53 of them were accepted for a full-text review, after which it was determined that only 36 papers contained relevant data for the objectives of this research. The summary of this review process is presented in Figure 3.

The main reasons for exclusion were that the analyzed papers did not contain information about modeling or CIMs. In three cases, the full text of the articles were not available.

Table 1 shows the publication date distribution of the selected papers. Note that 2013 only included the period between January and August of that year.

Analysis of the Indicators Collected from the Selected Papers
Table 2 details a summary of the information collected in the paper review. The complete list of publications and information collected and can be found in the supplementary material.

Of the selected papers, 50% focused on one specialized care department, while the others focused on multiple departments, national/regional projects, or described a theoretical approach. The papers cover a large variety of clinical domains, including nursing, oncology, neonatology, genetics, and infectious diseases. Most of the papers (83.3%) described a real-world deployment of CIMs. In addition, 73.2% of the papers mentioned the participation of health professionals during the CIM development process.

The preferred type of technical artefacts used to implement CIMs were archetypes (44.4%), followed by HL7 templates (25.0%). With regard to the RM used for the definition of CIMs, openEHR (25.0%), HL7 v3 (25%), including messages and CDA, and EN ISO 13606 (16.7%) were most frequently mentioned. Other works made use of proprietary RMs, expressed in UML, XML, or as ontologies.

All the references included in this systematic review apply a CIMP for defining CIMs, but only 52.8% of them described the CIMP they apply with some degree of detail.

In most of the studied papers, modeling of CIMs was centered on structural definition (e.g., a hierarchy of fields and grouping headings) without detailing how these structures were bound to terminologies (i.e., without mapping the field names to terms or specifying
terminology value lists for fields with textual values). Of the analyzed papers, 36.1% did not include any mention of the use of terminologies. In the others, SNOMED CT was the most widely adopted terminology (22.2%). Only four of the papers provided a detailed description of how they conducted the terminology binding process. The studies also lacked information on the metadata associated with the CIMs created (provenance, authorship, endorsements, related bibliography, etc.).

Publicly sharing the defined CIMs at the end of the CIMP is a means of providing credibility for and encouraging acceptance of developed artefacts and to facilitate their reuse. Only 38.9% of papers mentioned sharing the defined CIMs publicly. However, 72.2% of papers mentioned reusing existing CIMs as part of the development process.

A recurring (and current) demand in healthcare is for the use and production of specific tools and processes to solve problems related to electronically recording clinical data. Using appropriate design tools helps users manage the complexity of a detailed specification and helps ensure the resulting model’s syntactical correctness. Tool use should, therefore, contribute to the quality of the CIMs. In this context, we found that 67.7% of publications mentioned the use of specific tools for the creation of CIMs. Archetype editors were most frequently used (41.6%), followed by UML or similar visual design tools (13.9%). The other papers mentioned using tools such as spreadsheets, mind maps, XML editors, or Protégé.

**Inductive Analysis of Clinical Information Modeling Processes**

**Steps**

After tagging and categorizing the extracted information on the CIMPs described in the selected papers, we found the following nonmutually
Table 2: Indicators of the analyzed papers

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Values</th>
<th>References</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of CIM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HL7 templates</td>
<td>[23–31]</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>EN ISO 13606 or openEHR archetypes</td>
<td>[20, 21, 32–45]</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>[1, 22, 46–54]</td>
<td>11</td>
</tr>
<tr>
<td>Reference Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HL7 v3/HL7 CDA</td>
<td>[23–26, 28, 29, 31, 36]</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>openEHR</td>
<td>[32, 33, 38, 39, 40, 41, 42, 44, 45]</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>EN ISO 13606</td>
<td>[20, 21, 34, 35, 37, 43]</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>[1, 22, 27, 46–54]</td>
<td>12</td>
</tr>
<tr>
<td>CIMP Is Described</td>
<td>Yes</td>
<td>[20, 21, 23, 26, 29, 31–35, 41, 43 45–49, 54]</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>[1, 22, 24, 25, 27, 28, 36–38, 40, 42, 44, 48, 50–53, 1, 22]</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>[24, 28, 29, 32, 44, 46, 47, 49]</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Not Specified</td>
<td>[48, 50]</td>
<td>2</td>
</tr>
<tr>
<td>Are resulting CIMs shared?</td>
<td>Yes</td>
<td>[30, 35–38, 41, 45, 53]</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Planned</td>
<td>[1, 20–22, 39, 40]</td>
<td>6</td>
</tr>
<tr>
<td>Terminologies Used</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNOMED CT</td>
<td>[24, 25, 33, 38, 41, 50, 52, 53]</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>[23, 28–32, 34, 35, 43, 46–49, 51]</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Not Specified</td>
<td>[1, 20–22, 26, 27, 36, 37, 39, 40, 42, 44, 45, 54]</td>
<td>14</td>
</tr>
</tbody>
</table>

(Continued)
exclusive categories of related information. Table 3 summarizes the papers, including information related to each category described below.

- **Scope definition leading to selection of the domain and selecting relevant experts.** Whether the scope of a CIM is local or designed for wider use, it will be needed to identify the domain to be covered and the expected uses of the CIMs to be developed. Based on the identification of the care setting, healthcare activities, and clinical requirements, it is possible to create a work group of relevant experts in that clinical domain who will be responsible for designing the CIMs.

- **Analysis of the information covered in the specific domain.** In order to create complete CIMs definitions, it is required to obtain an understanding of clinical scenarios, workflows, and users, to determine the data items that will be supported by CIMs. It is necessary to identify how the existing systems have been implemented and documented. Reviewing guidelines, literature, and validated clinical scales allows the design team to ensure that information covered by the CIMs will meet the requirements of clinical practice. To collect this information, interviews and workshops with clinical experts may be performed.

- **Design of CIMs.** After identifying the necessary data items, they are merged and harmonized into CIMs, avoiding possible overlapping. Each CIM will detail the possible set of attributes associated with it in a structured way. Each data item associated with a clinical concept can be detailed in the form of a value set or CIM node. It is also important to identify domain terminologies that are applicable to the studied domain, in order to map them to the CIMs. The definition of CIMs can be focused either on just determining the essential data
sets as common minimum communication requirements or on satisfying the application of CIMs for multiple purposes, ensuring a basic compatibility across domains.

- Definition of implementable CIM specifications. In order to make the defined CIMs compatible with existing EHR information standards, implementable technical specification is needed. The process of implementing the modeled CIMs into technical artifacts starts with the search and review of existing CIMs. Those CIMs that suit the scope of the project will be reused or adapted. This will increase the interoperability between systems with different local needs, using similar CIMs. For those clinical concepts that are not covered by existing CIMs, new ones will be created.

- Validation. Multiple techniques have been adopted to validate the defined models, including peer review and the creation of prototype screens. Further evaluation, using routinely collected clinical data from multiple patients, will provide stronger validation for the defined CIMs.

- Publishing and maintenance. Those CIMs that are created should be transferred into a public repository in order to be accessible by any other user. CIMs published in the repository should include a method for receiving feedback from those projects and organizations that adopt them.

- Governance. This final category is not properly a step of the CIMP but is closely related to it. The organization responsible for developing and maintaining CIMs will be in charge of establishing effective governance of them. This governance will determine the process for quality review and publication of CIMs and will also maintain relationships with other projects and organizations working in the same domain covered by those CIMs.

**Discussion on the Extracted Indicators**

Indicators extracted from the selected papers raise several interesting discussion points.

- Limited information about the CIMP used to create CIMs. All the selected papers relied on the use of CIMs as a key piece of their information systems. However, the methodology followed to create these CIMs was not usually described in detail and sometimes not mentioned at all. This lack of information might reduce the level of third-party trust in the quality of the developed CIMs. Given that, currently, a standard CIMP does not exist in the literature, we had expected that more authors would have included a detailed description of their own modeling and validation steps. We noted that information on terminology bindings provided in the studied papers was particularly limited. Thirty-six percent of the reviewed papers did not even mention the terminological aspect, and most of the others only referred to it as a part of the work to be completed in the future. A CIM cannot be semantically interoperable if it lacks terminological references that describe its contents. The definition of a particular information structure can be affected by the expressivity of the selected terminologies that accompany it and, vice versa, the design of a particular information structure affects how the value sets to be used in it should be created. Moreover, a loose definition and use of terminological value sets also affects the final quality and interoperability of the clinical data produced.

- Resultant CIMs are not shared. It was observed that most of the analyzed experiences did not provide any mechanism to access the resulting CIMs. Although it is not mandatory to share CIMs with external groups, it would be a good practice to share these models openly (unless there are copyright restrictions). This could improve the quality of the defined models through feedback and support the harmonization of multiple groups developing CIMs in parallel in the same domain (and, thus, the semantic interoperability of EHR information).

- Modeling tools. CIM definition is a multidisciplinary process in which health professionals and technicians collaborate. To this end, it is important to have the appropriate tools, to ease the definition and review processes. This study suggests that most modeling efforts use generic tools to carry out this work, such as UML technologies, mind maps, spreadsheets, or XML tools.
Only those that rely on the archetype approach make use of specific tools. In any case, several of the reviewed papers warn of the immaturity of modeling support tools. We can conclude from these results that there is a need for better modeling tools. However, it has to be taken into account that the papers that noted a lack of sophisticated modeling tools are from 2007, 2009, and 2011. Most likely, improvements have been made in this area over the last few years.

- **Mapping to implementable specifications.** Transforming generic CIMs definitions into implementable specifications (ie, archetypes or templates) is not a direct process, because it requires accommodating the information attributes of the CIM in a specific RM structure. This implies that a shared CIM could be implemented in different technical artefacts or in standards that are not completely equivalent.

**Discussion on the Inductive Analysis Results**

The methodological approach to create CIMs was found to be similar in all the studied papers in which information on the approach was available. Figure 4 summarizes the steps obtained from the inductive analysis of the content related to CIMP and the relationships between them. The process starts with selecting the scope and the work team, followed by a domain analysis (which includes researching references or existing CIMs that could be reused), designing and defining the structure and semantics of new CIMs (or the modification of existing ones), validation by health professionals, and, finally, publishing the resulting CIMs.

Although these steps were defined based on the partial information available in the published literature, the level of similarities found suggests that it would be possible to define a unified process to guide CIMs definition, including a description of best practices to increase the quality of the CIMs.

Finally, the identified CIMP steps are encompassed by a general governance process. This governance is responsible for identifying when a new CIM needs to be created and when existing ones should be reviewed. The governance of CIMs is a separate topic that has also received attention by researchers.

**Limitations and Risk of Bias of This Systematic Review**

The authors recognize that including the “semantic interoperability” criterion could have limited the papers found in the search, since use of this term in the early 2000s was limited. However, including this criterion allowed us to collect early experiences of CIM-based approaches from early promoters of the semantic interoperability concept. Nearly 20 papers published before 2005 that included the term were found.

In order to limit the risk of bias in this systematic review, all papers were screened by at least two of the authors of this paper, who had to agree on their suitability for the full-text review phase. In the full review phase, the authors’ roles for the papers switched. Every paper was screened by one set of authors and fully reviewed (if it was determined to be suitable for Phase 2 of the review) by a different author. In the inductive analysis, the authors also achieved a consensus on the steps and classifications of the selected papers, based on the information contained in them.

Regarding the obtained results and conclusions, one limitation of the review we performed is that most papers did not describe in detail the CIMP that was followed in order to define the CIMs. In many cases the modeling process was mentioned once or only briefly throughout the text. This necessitated a careful and detailed reading.
of each of the papers in order to determine the CIMP steps followed by the original authors.

CONCLUSION
The use of CIMs has gained recognition as one of the essential aspects of the creation of standardized and interoperable EHR systems. Different standards and technical approaches exist (eg, EN ISO 13606 and openEHR, using archetypes, or HL7 v3, using templates), but all of these methodologies share the idea of separating the definition of the CIMs from the actual representation and persistence of the data values. Moreover, the work of international modeling initiatives, such as the Clinical Information Modeling Initiative, indicates an increased interest in creating reusable CIMs. Thus, it is important that the CIMP used to create those models follows clear and well-defined steps.

This research characterized published experiences related to the creation of semantically interoperable EHR systems between 2000 and 2013, in order to obtain a better understanding of the steps followed during the creation of CIMs. It was found that most of the articles reported using a similar approach for CIMs creation. This suggests that it should be possible to create a common or unified methodology for future clinical information modeling. This conclusion is, however, limited, due to the selected papers’ lack of detail on the CIMs used. It is important to advocate for further collaboration between the main organizations and professionals involved in CIM development, to reach a consensus on a unified best practice CIMP.

A commonly agreed upon CIMP will promote and emphasize the importance of analyzing the information covered in a particular domain, the collaboration between different clinical and technical professionals, and the search for consensus in the definition of CIMs. It will also minimize the diversity of ways in which a CIM can be designed and will make terminology bindings more consistent. This is directly related to the improvement of the quality of CIMs.59,60

CONTRIBUTORS
A.M., D.M., W.D.C., and M.R.S. conceived and designed the study and performed the systematic review. A.M. and D.M. have been the main contributors to the manuscript. J.A.M., M.R., and D.K. have critically revised the manuscript and provided insights on the review discussion and conclusions. All authors approved the final manuscript.

FUNDING
This research has been partially funded by the Instituto de Salud Carlos III (Platform for Innovation in Medical Technologies and Health), grant PT13/0006/0036 and the Spanish Ministry of Economy and Competitiveness, grants TIN2010-21388-C02-01 and PTQ-12-05620.

COMPETING INTERESTS
None.

SUPPLEMENTARY MATERIAL
Supplementary material is available online at http://jamia.oxfordjournals.org/.

REFERENCES