Paper

An Ontological Framework for Representing Clinical Knowledge in Decision Support Systems

Marco Iannaccone and Massimo Esposito

National Research Council of Italy, Institute for High Performance Computing and Networking, Naples, Italy

Abstract—In the last decades, clinical evidence and expert consensus have been encoded into advanced Decision Support Systems (DSSs) in order to promote a better integration into the clinical workflow and facilitate the automatic provision of patient specific advice at the time and place where decisions are made. However, clinical knowledge, typically expressed as unstructured and free text guidelines, requires to be encoded into a computer interpretable form suitable for being interpreted and processed by DSSs. For this reason, this paper proposes an ontological framework, which enables the encoding of clinical guidelines from text to a formal representation, in order to allow querying, advanced reasoning and management in a well defined and rigorous way. In particular, it jointly manages declarative and procedural aspects of a standards based verifiable guideline model, named GLM-CDS (GuideLine Model for Clinical Decision Support), and expresses reasoning tasks that exploit such a represented knowledge in order to formalize integrity constraints, business rules and complex inference rules.

Keywords—Clinical Practice Guidelines, Decision Support Systems, Ontology, Rules, Unstructured Data.

1. Introduction

In the last years, healthcare has been more and more characterized by an extensive practice variation and overuse, underuse, and misuse of medical resources. To address these issues, both clinical evidence and expert consensus have been systematically captured and joined to encode Clinical Practice Guidelines (CPGs), aimed at supporting general practitioners in making clinical decisions and managing medical actions about appropriate healthcare for specific clinical circumstances [1], [2].

Most CPGs, however, are expressed in a text-based format and, thus, are not easily accessible to care providers, who need to apply them either at the time and place where clinical decisions are made, or to assess the quality of their application, retrospectively.

Even if, recently, CPGs have been published also in electronic formats, such as HTML or PDF files, they are poorly adopted and examined by care providers [3], who rarely have the time to utilize the valuable knowledge, encoded in the guidelines, during the treatment of their patients. Therefore, there is a need to facilitate automated guideline specification, dissemination, application, and quality assessment in order to realize the actual potential of CPGs in improving health outcomes.

Several recent studies have suggested that automation might be realized by encoding CPGs into advanced Decision Support Systems (DSSs), i.e., computer-based systems designed to promote a better integration into the clinical workflow and to facilitate the automatic provision of patient-specific advice at the time and place where decisions are made [2], [4]. However, this requires clinical knowledge, expressed into an unstructured and free-text format, to be encoded as computer-interpretable guidelines (CIGs), suitable for being interpreted and processed by DSSs.

Even if, in the recent past, many knowledge representation formalisms have been developed to address this issue, it remains strongly critical, since a mismatch exists between the unstructured narrative form of published CPGs and the formality that is necessary for the operationalization of clinical knowledge in CIGs for DSSs. Moreover, the poverty of the methodological rigor typically used to computerize guideline knowledge further complicates this operationalization, which might generate malformed, incomplete, or even inconsistent CIGs.

For this reason, this paper proposes an ontological framework, which enables the encoding of CPGs from text to a formal representation, where domain knowledge, clinical process structures and data, and the behavioral semantics of such processes are encoded in order to allow querying, advanced reasoning and management in a well-defined and rigorous way.

In particular, it jointly manages declarative and procedural aspects of a standards-based verifiable guideline model, named GuideLine Model for Clinical Decision Support (GLM-CDS) [5], and expresses reasoning tasks that exploit such a represented knowledge in order to formalize integrity constraints, business rules and complex inference rules.

The solution here proposed is particularly relevant for the design and development of a CIG, enabling the possibility of inferring implicit knowledge not expressly formulated or verifying the consistency and coherency of the knowledge explicitly modeled.

The rest of the paper is organized as follows. Section 2 outlines an overview of the state-of-the-art solutions. In Section 3, the proposed framework is described by referring to the guideline model used, i.e. GLM-CDS. Section 4

depicts an example application in order to highlight how the proposed framework can be used to formalize an existing guideline defined in GLM-CDS. Finally, Section 5 concludes the work.

2. Related Work

To date, the effort in defining new solutions for computerizing CPGs has produced many process-flow-like models, such as SAGE [6], GLIF [7], Asbru [8], EON [9] and PROforma [10], which are characterized by different coverage and particularities in order to represent both the structure of the domain-specific knowledge, named declarative knowledge, and the process-oriented knowledge, named procedural knowledge.

In particular, declarative knowledge concerns the domain compositional elements, such as raw and abstract concepts, their properties and inter-relations explicitly expressed in the CPGs.

On the other hand, procedural knowledge captures the control-flow logic to be modelled by providing suggestions about the actions to be taken or conclusions to be drawn from *declarative knowledge*, as well as constraints between tasks, temporal constraints in a global plan, and so on [11].

All the above-mentioned models are process-flow-like and share same basic procedural elements: some kind of action/decision tasks, some implicit or explicit mechanisms for coordination or synchronicity of actions, the ability to create sub-plans or sub-guidelines, the possibility of storing the state of a guideline which is being executed and synchronizing the management of a patient with the corresponding parts of a guideline by means of some entry/exit points [11].

Various types of actions can be supported, such as medically-oriented (e.g., recommending the administration of a particular substance) or programming-oriented (e.g., notifying a message to a care provider).

Moreover, two basic types of decisions are mainly defined: decisions in the form of if-then-else choices and decisions requiring a heuristic choice from a set of rule-in and rule-out conditions that support or oppose alternatives [12].

A drawback common to all these proposals is represented by the lack of a seamless integration of both declarative and procedural knowledge expressed in a CPG by means of a highly expressive and formal framework able to jointly manage control-flow and domain-specific aspects and express reasoning tasks to automatically infer implicit knowledge or verify a number of desired properties of correctness, coherency and well-formedness of a CIG, also with respect to the time perspective.

The solution here proposed has been conceived to face these issues by expressing in a combined way domain ontologies, clinical processes, related decision and inference rules, and integrity constraints, as described in the following sections.

3. The Ontological Framework for Computerizing CPGs

The formal framework here proposed is aimed at computerizing CPGs by defining a guideline model, named GLM-CDS, and by encoding such a model through a hybridization of the theoretic semantics of ontology and rule languages.

Deeply speaking, the proposed model GLM-CDS consists of a control-flow part, which is based on a formal Task-Network Model (TNM) for codifying CPGs in terms of structured tasks connected with transition dependencies between them from an initial state of the patient.

Domain-specific knowledge is coded through an information model built on the top of the Domain Analysis Model, Release 1 of the HL7 Virtual Medical Record [13] (HL7 vMR-DAM) issued by HL7 Clinical Decision Support-Working Group. This information model is populated by using existing standard terminological resources, such as Logical Observation Identifiers Names and Codes (LOINC) [14] and Systematized NOmenclature of MEDicine (SNOMED) [15].

Data types used in GLM-CDS resemble the ones defined in the HL7 vMR DAM, which gives a simplified/constrained version of ISO 21090 data types, based on the abstract HL7 version 3 data types specification, release 2 [16].

The control flow part is formally defined as the following 8-tuple:

$$C_f = \langle G, E_n, E_x, T, C, D_r, I_r, C_s \rangle \,, \tag{1}$$

where:

- G indicates the set of *sub-guidelines* included into a CPG,
- E_n and E_x represent the *entry point* and the *exit point* of the TNM modeling a CPG,
- T represents the set of *tasks* composing a CPG,
- C is the set of *connections* between the nodes of a TNM,
- D_r is the set of *decision rules*, which relate a decision node to a task node and are used at runtime to automatically control the execution flow of a process,
- I_r is the set of *inference rules*, which combine known knowledge to produce ("infer") new information,
- C_s is the set of *constraints*, which have to be verified in order to preserve the correctness, coherency and consistency of the CPG modeled.

Moreover, T is partitioned into the following sub-sets:

- D is the set of *decision* nodes for directing the control-flow from a point into the TNM to various alternatives,

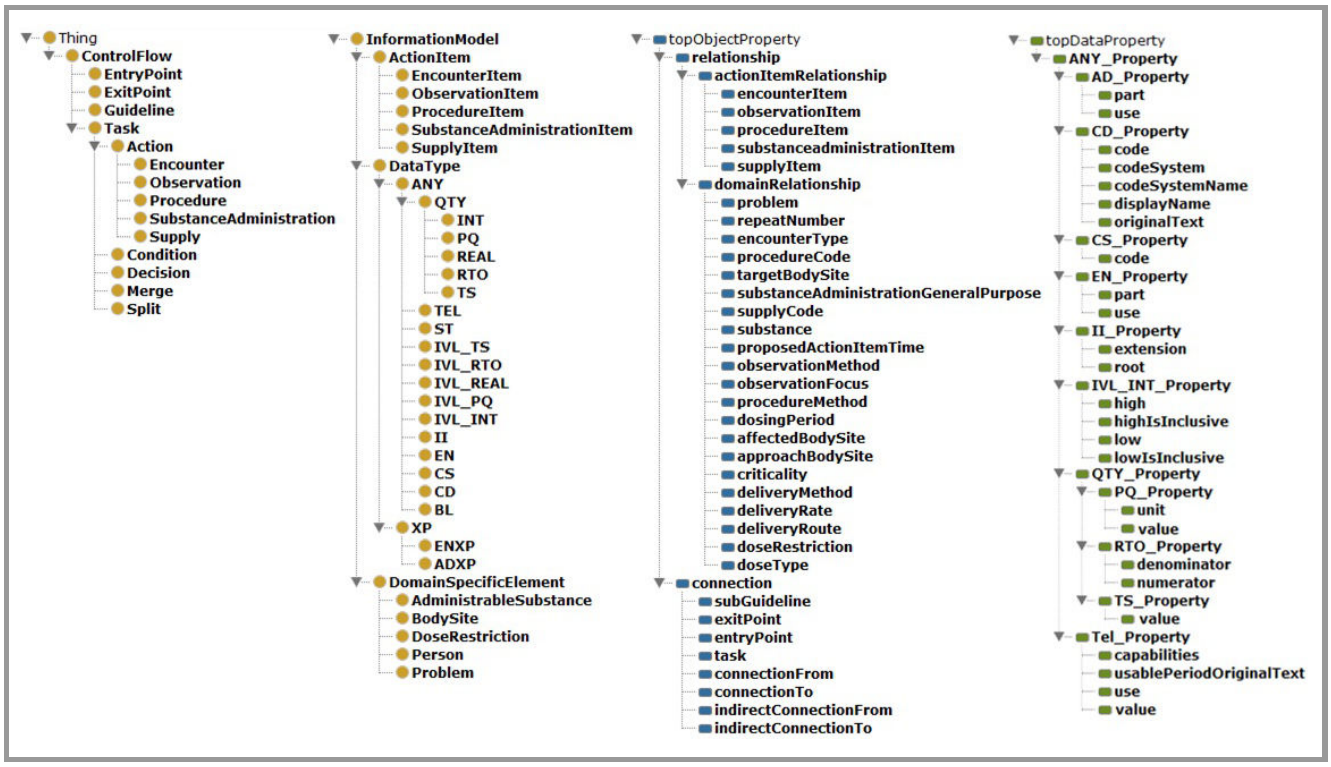

Fig. 1. A compacted perspective of both ontology concepts and roles formalized in GLM-CDS.

- C_n is the set of *conditions*, defined as observable states of the patient that persist over time and tend to require intervention or management,
- S indicates the set of *split* nodes, which enable to branch the guideline flow to multiple parallel tasks,
- M indicates the set of *merge* nodes, which enable to synchronize parallel tasks by making them converging into a single point,
- A models the set of high level *actions* to be performed and is further specialized into the following sub-sets:
  - A_O is the set of *observations*, which are used to determine a measurement, a laboratory test or a user input value,
  - A_S models the set of *supplies*, which are aimed at providing some clinical material or equipment to a patient,
  - A_E is the set of *encounters*, which are applied to request an appointment between a patient and healthcare participants for assessing his health status,
  - A_P represents the set of *procedures*, whose outcome is the alteration of the patient's physical condition,
  - A_{SA} refers to the set of *substance administrations*, which allow giving a substance to a patient for enabling a clinical effect.

Furthermore, C is also composed by the following sub-sets:

- C_d indicates the set of *direct connections* between pairs of nodes of a TNM without other intermediary nodes,
- C_i indicates the set of *indirect connections* between pairs of nodes of a TNM with other intermediary nodes.

Finally, C_s is partitioned into:

- IC_s indicates the set of *integrity constraints* devised to detect violations, errors and/or missing information in the TNM encoding a CPG,
- TC_s represents the set of *temporal constraints* formulated according to some time patterns, i.e. task duration, periodicity, deadline, scheduling and time lags.

On the other hand, the information model is formalized as the following 5-tuple:

$$I_m = \langle A_I, E_D, R_D, P_D, D_T \rangle \,, \tag{2}$$

where:

- A_I models the set of elementary and repeatable *action items*, associated to each action and specialized into:
  - I_O the set of *observation items*,
  - I_S the set of *supply items*,
  - I_E the set of *encounter items*,
  - I_P the set of *procedure items*,
  - I_{SA} the set of *substance administrations*.

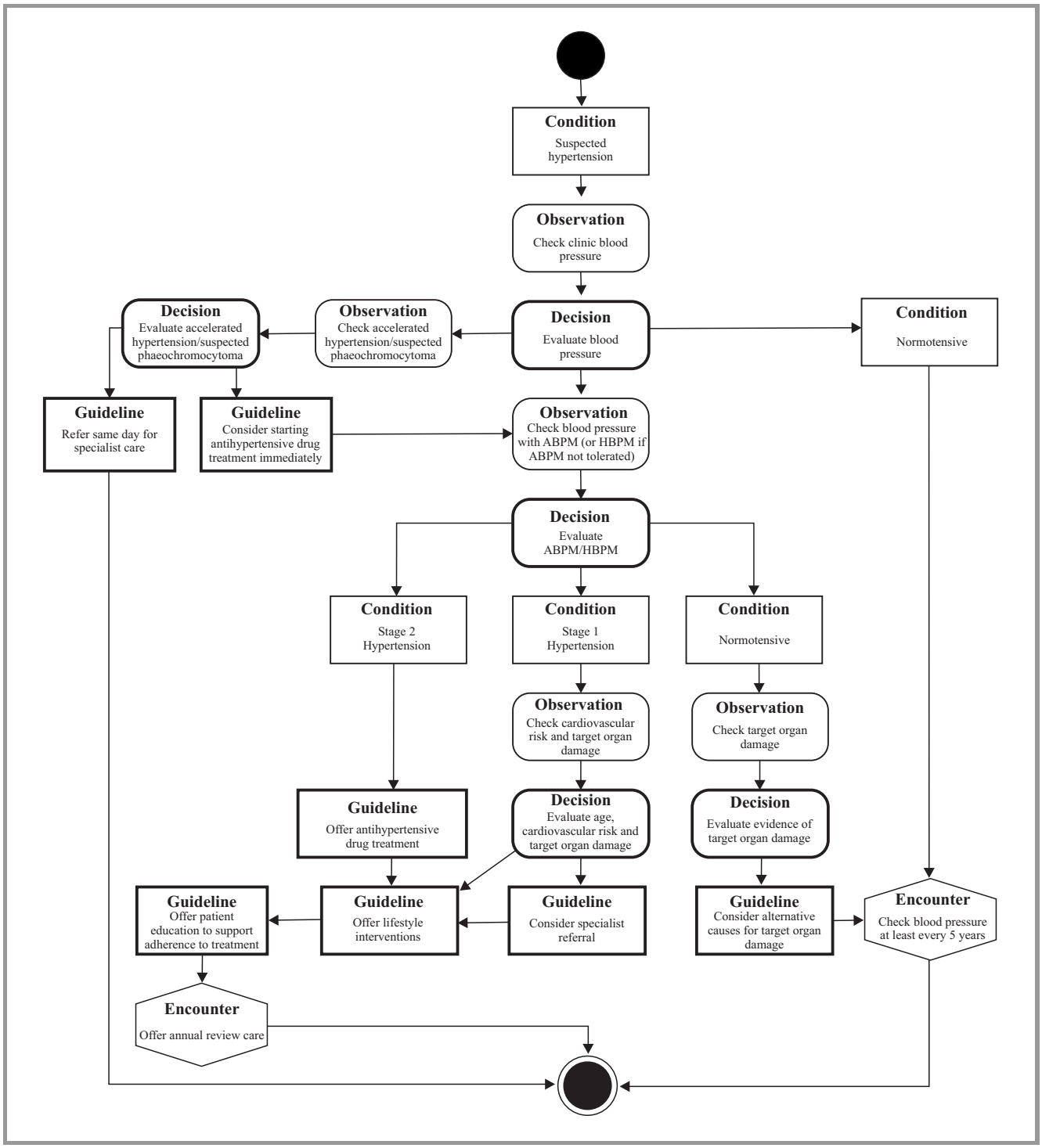

Fig. 2. A fragment of the NICE guideline for hypertension in adults encoded in GLM-CDS.

- E_D models the set of *domain-specific elements*, such as *Administrable Substance, Dose Restriction* or *Body Site*, which are linkable to the action items,
- R_D indicates the set of relationships existing between:
  - *action items* and *domain specific* concepts,
  - *action items* and *data types*,
  - *domain-specific concepts* and *data types*,
  - elements belonging to the subset $G \cup E_n \cup E_x \cup T$ of the control flow part and *data types*.
- P_D represents the set of properties used for specifying values a data type can assume,
- D_T models the set of data types used in GLM-CDS.

This guideline model has been encoded by exploiting the theoretic semantics of ontology and rule languages.

In detail, ontology languages rely on decidable fragments of first order logic and are based on the notions of concepts (unary predicates, classes), individuals (instances of concepts), abstract roles (binary predicates between concepts) and concrete roles (binary predicates between concepts and data values).

On the other hand, rule languages are widely considered in literature as a syntactic and semantic extension to ontology languages. Indeed, rules have been widely used as a new kind of axiom to define abstract roles as well as arithmetic relationships between data values assumed by concrete roles.

As a result, a subset of the control flow part C_f of GLM-CDS, i.e. $G \cup E_n \cup E_x \cup T$, as well as a subset of its information model I_m, i.e. $A_I \cup E_D \cup D_T$, have been encoded as ontology concepts. Furthermore, the sets C of C_f and R_D of I_m have been encoded by means of ontology abstract roles, whereas the set P_D of I_m has been formalized by using ontology concrete roles.

Figure 1 reports a compacted perspective of ontology concepts and abstract roles formalized in GLM-CDS.

The sets D_r and I_r of decision and inference rules are formulated by using the Horn Clause Logic. In particular, decision and inference rules are expressed as definite Horn clauses, in the form:

$$h_1(X_1) \leftarrow b_1(Y_1) \wedge \cdots \wedge b_k(Y_k), \tag{3}$$

where the clause $h_1(X_1)$ is named head, the clauses $b_1(Y_1) \ldots b_k(Y_k)$ (with $k \geq 0$) are called *body*, $h_1, b_1 \ldots b_k$ are rule predicates and $X_1, Y_1 \ldots Y_k$ are tuples of variables or constants. Rule predicates are built by using ontology concepts and roles and by using ontology individuals as constants. Moreover, each variable in the head is obliged to appear also in the body of a rule, so granting soundness and completeness of the reasoning process.

Finally, the sets IC_s and TC_s of integrity and temporal constraints are expressed as negative Horn clauses, in the form:

$$\leftarrow b_1(Y_1) \wedge \cdots \wedge b_k(Y_k), \tag{4}$$

where no clause is reported in the head. For the sake of uniformity with decision and inference rules, each constraint is associated with a special predicate C_s , which indicates whether it is violated, as formulated in (5):

$$C_s \leftarrow b_1(Y_1) \wedge \cdots \wedge b_k(Y_k). \tag{5}$$

4. An Example Application: a CIG for Hypertension in Adults

This section reports, as an example, the application of GLM-CDS to the CPG for the "Clinical management of primary hypertension in adults", issued by the National Institute for Health and Care Excellence (NICE).

A fragment of this CPG, containing recommendations on blood pressure measurement, the use of ambulatory/home blood pressure monitoring (ABPM/HBPM) and the management of hypertension, has been encoded according to the GLM-CDS as shown in Fig. 2.

A partial translation of this fragment of CPG into its ontological representation is given in Fig. 3.

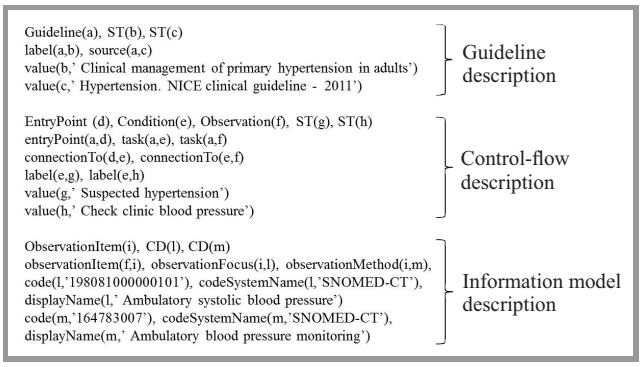

Fig. 3. A partial translation of a fragment of the guideline into its ontological representation.

In detail, the CPG is first described, by listing some relevant information, such as its name and source. Next, the first three nodes, i.e. the Entry Point, the Condition named "Suspected hypertension" and the Observation named "Check clinic blood pressure", of the control-flow part reported in Fig. 2 are formalized as ontology concepts and roles.

Finally, with respect to Observation named "Check clinic blood pressure", a specific Observation Item is codified, whose roles *observationFocus* and *observationMethod* are valued according to the data type *CD* of the HL7 vMR-DAM.

An example of decision rule, associated to the Decision node named "Evaluate blood pressure", is reported in Fig. 4. It evaluates whether the systolic blood pressure is less than 140 mmHg.

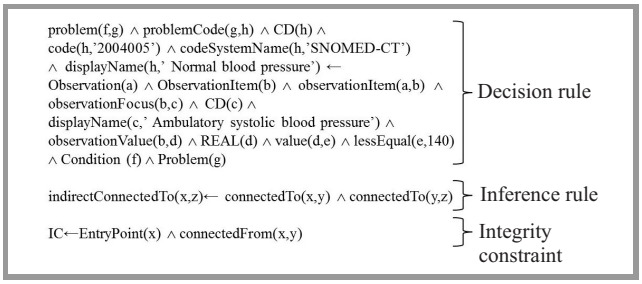

Fig. 4. Some examples of decision rules, inference rules and integrity constraints.

Moreover, an inference rule, which defines the role *indirectConnectionTo* starting from the role *connectionTo*, is also reported.

Finally, an example of integrity constraint is also codified, which states that each Entry Point node of a CPG is not admissible to have a direct connection from any other node.

5. Conclusion

To date, the different attempts proposed for encoding CPGs in a computer interpretable form suitable for DSSs are not fully concerned with enabling an intuitive and, contextually, formal representation of CPGs, in terms of their logic, the clinical processes involved and the different types of clinical knowledge represented.

For this reason, this paper proposed an ontological framework for encoding CPGs from text to a formal representation, by jointly managing declarative and procedural aspects of a standards based verifiable guideline model, named GLM-CDS, and expressing reasoning tasks that exploit such a represented knowledge in order to formalize integrity and temporal constraints, business rules and complex inference rules.

The strength of this solution relies on the support to design and develop a CIG, by enabling the possibility of inferring implicit knowledge not expressly formulated or verifying the consistency and coherency of the knowledge explicitly modelled.

In order to promote and facilitate the widespread use of this framework, ongoing activities are being carried out to design and realize an ad hoc, intuitive and user friendly authoring tool for encoding CPGs graphically and, successively, translating them into a formal representation expressed in terms of ontology concepts and roles, as well as decision/inference rules and integrity/temporal constraints.

Acknowledgements

This work has been partially supported by the Italian project "ASK-Health" Advanced system for the interpretation and sharing of knowledge in the healthcare sector.

References

[1] B. S. Bloom, "Crossing the quality chasm: a new health system for the 21st century", *JAMA: The J. Amer. Medical Assoc.*, vol. 287, no. 5, pp. 646–647, 2002.

[2] M. J. Field and K. N. Lohr, *Guidelines for Clinical Practice: From Development to Use*. Washington: National Academy Press, 1992.

[3] F. A. Sonnenberg and C. G. Hagerty, "Computer-interpretable clinical practice guidelines: Where are we and where are we going?", *Methods Infor. Med.*, vol. 45, no. 1, pp. 145–158, 2006.

[4] F. Moscato, V. Vittorini, F. Amato, A. Mazzeo, and N. Mazzocca, "Solution workflows for model-based analysis of complex systems". *IEEE Trans. Autom. Sci. Engin.*, vol. 9, no. 1, pp. 83–95, 2012.

[5] M. Iannaccone, M. Esposito, and G. De Pietro, "A standards-based verifiable guideline model for decision support in clinical applications", in *Process Support and Knowledge Representation in Health Care*, D. Riano, R. Lenz, S. Miksch, M. Peleg, M. Reichert, and A. Teije, Eds. Springer, 2013, pp. 143–157.

[6] S. W. Tu *et al.*, "The SAGE guideline model: Achievements and overview", *J. Am. Med. Inform. Assoc.*, vol. 14, pp. 589–598, 2007.

[7] A. A. Boxwala *et al.*, "GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines", *J. Biomed. Inform.*, vol. 37, no. 3, pp. 147–161, 2004.

[8] A. Seyfang, S. Miksch, and M. Marcos, "Combining diagnosis and treatment using Asbru", *Int. J. Med. Inform.*, vol. 68, no. 1, pp. 49–57, 2002.

[9] S. W. Tu and M. A. Musen, "Modeling data and knowledge in the EON guideline architecture", *Stud. Health. Technol. Inform.*, vol. 84, no. 1, pp. 280–284, 2001.

[10] J. Fox and N. Johns, "Rahmanzadeh A. Disseminating medical knowledge: the PROforma approach", *Artif. Intell. Med.*, vol. 14, pp. 157–181, 1998.

[11] D. Isern and A. Moreno, "Computer-based execution of clinical guidelines: a review", *Int. J. Med. Inform.*, vol. 77, pp. 787–808, 2008.

[12] P. De Clercq, K. Kaiser, and A. Hasman, "Computer-interpretable guideline formalisms", *Stud. Health. Technol. Inform.*, vol. 139, pp. 22–43, 2008.

[13] "HL7 Virtual Medical Record (vMR) Project Wiki", Health Level 7 [Online]. Available: http://wiki.hl7.org/index.php?title=Virtual_Medical_Record_(vMR)

[14] "Logical Observation Identifiers Names and Codes (LOINC)", Regenstrief Institute, Inc and the LOINC Committee [Online]. Available: http://loinc.org/

[15] "Systematized Nomenclature of Medicine (SNOMED)", International Health Terminology Standards Development Organisation [Online]. Available: http://www.ihtsdo.org/snomed-ct/

[16] "HL7 Reference Information Model, Version 3", Health Level 7 [Online]. Available: http://www.hl7.org/implement/standards/rim.cfm

Marco Iannaccone is a research assistant at the Institute for High Performance Computing and Networking (ICAR) of the National Research Council of Italy (CNR). He received his M.Sc. in Computer Science Engineering from University of Naples Federico II in 2004. Since July 2012, he has been a member of the iHealthLab – Intelligent Healthcare laboratory. His research interests cover knowledge representation and ontologies for healthcare processes, process modeling, interoperability and standards in healthcare, workflow management in healthcare. His research is described in scientific articles published in international conferences.

E-mail: marco.iannaccone@na.icar.cnr.it
National Research Council of Italy (CNR)
Institute for High Performance Computing
and Networking (ICAR)
Via Pietro Castellino, 111
80131, Naples, Italy

Massimo Esposito is a scientific researcher at the Institute for High Performance Computing and Networking (ICAR) of the National Research Council of Italy (CNR). He received his M.Sc. in Computer Science Engineering (Cum Laude) from University of Naples Federico II in 2004. He received a University 1st level Master

degree, named European Master on Critical Networked Systems in 2007, and a Ph.D. degree in Information Technology Engineering in 2011 from the University of Naples Parthenope. Since 2007, he has been a member of the Advanced Medical Imaging and Computing labOratory (AMICO), developed from a cooperation agreement between the Institute of Biostructure and Bioimaging (IBB) and ICAR of CNR. Since 2011, he has been a member of the iHealthLab – Intelligent Healthcare laboratory. His research interests cover pervasive computing, knowledge-based medical decision support systems, knowledge discovery in biomedical databases, workflow management in healthcare. His research is described in many scientific articles published in international conferences and journals. He participates in the editorial boards of many international journals and has been on the program committee of many international conferences.

E-mail: massimo.esposito@na.icar.cnr.it
National Research Council of Italy (CNR)
Institute for High Performance Computing
and Networking (ICAR)
Via Pietro Castellino, 111
80131, Naples, Italy