Distributed Guidelines (DiG): A Software Framework for Extending Automated Health Decision Support to the General Population

by Edwin Yaqub, MSc, and Andre Barroso, PhD

Abstract

This article presents a framework and methodology to create personal health record (PHR) systems able to transform raw health data into meaningful information for the general population. By bridging the semantic gap between an individual and his or her health data, it is expected that better care will ensue through consumer empowerment. An important challenge for the realization of this vision is the lack of available expert knowledge in a format that is concomitantly easy to codify, share, and be used by the general population. To address this challenge, we developed a novel approach to encode expert knowledge into machine-interpretable, reusable components called "consumer guidelines." Once encoded, guidelines are easily shared, extended, and modified. These guidelines can exist as distributed documents on the Internet and be executed by our processing engine (Health-Guru) to provide an individual with personalized assessment against various health risks based on the evidence data stored in a PHR.

Key words: personal health records; decision support techniques; semantic web; ontology; public health; clinical guidelines

Introduction

From the patient's perspective, the usefulness of personal health record (PHR) systems is intrinsically related to their ability to combine health data and share relevant data with professionals for comprehensive analysis.[1, 2] In combination with home sensors used for balanced-lifestyle maintenance, disease prevention, or disease management, such systems may add great value by helping patients to directly interpret the obtained data in the context of their overall health status. In order to accomplish this goal, a certain degree of expert knowledge is needed.

Health information is often rich in technical jargon, making its interpretation by nonspecialists difficult. Access to information that helps patients to translate their data into actionable information is thus desirable. The first step in this direction is to bring data to a level of abstraction with which nonspecialists are familiar. The value and significance of this process is recognized by governmental agencies all over the world, and considerable amounts of money are

spent to educate and create awareness among the population as a means to reduce the social impact of diseases.

Expert knowledge for the general population is generally delivered through leaflets, TV or radio campaigns, and in recent years also Web sites.[3, 4] This information needs to be conveyed in simple, brief, and interesting ways to capture and maintain the target audience's attention. Unfortunately, not all relevant details can be properly represented in this manner. The very nature of the media makes the information susceptible to quick and short-lived consumption. On the other end of the complexity spectrum, expert knowledge is codified in specialized publications that carry fine levels of detail for completeness and accuracy.[5] These publications often impose on the patient a much greater interpretation effort and the burden of applying the generic information to his or her particular case.

Clinical guidelines represent another rich source of information that patients can utilize in order to improve their understanding of health issues.[6] Nonetheless, the usability of clinical guidelines by patients is somewhat compromised by the generic nature of their content: the reader still needs a certain degree of expert knowledge to apply the guidelines to concrete cases.

A system able to easily capture detailed expert knowledge and automatically combine it with personal health data as a means to deliver meaningful and actionable information for patients and the general population is therefore needed.

This article introduces Distributed Guidelines (DiG), a framework for building PHR systems and applications that serves as a tool for assisting patients to make sense of their health data in light of expert knowledge. The framework also facilitates the acquisition and dissemination of expert knowledge.

Background

Clinical guidelines are systematically developed statements that assist clinicians and patients in making decisions about appropriate treatment for specific conditions.[7] These statements are created based on best-practice evidence and aim to improve the quality of healthcare by making this information accessible to the population. Awareness of the benefits of clinical guidelines prompted governmental and nongovernmental agencies interested in public health to systematically compile and make guidelines available to the population.[8, 9] Despite the advantages and efforts to make clinical guidelines more accessible, studies indicate that many clinicians do not follow them in their daily practice.[10–12] Observed usage levels depend on a number of factors, including time pressure and the intricacy of efficiently applying all guideline steps in the daily routine of the practice.[13–15]

Computerized information and decision support systems are an important class of tools to promote the use of clinical guidelines among clinicians. Such systems can automatically synthesize patient data, perform complex evaluations, and generate specific recommendations in a timely manner. A typical clinical decision support system comprises the following building blocks (see Figure 1).

- **Electronic Health Record (EHR)**: stores and retrieves patient health information
- **Knowledge Base**: stores and retrieves domain or expert knowledge in a format that can be parsed and executed by computers
- **Execution Engine**: parses and executes guidelines stored in the knowledge base on behalf of clinical applications

- **Clinical Application**: interfaces with the user and integrates all other blocks

Over the past two decades, various computer-based systems were developed to integrate clinical guidelines into healthcare professionals' workflow (e.g., EON, GuideLine Interchange Format [GLIF], GASTON, and SAGE).[16-19] A comprehensive list of existing systems can be found on the OpenClinical Web site.[20]

Solutions for integrating patient-entered data into the workflow of clinical decision support systems have also been explored for many years, and activities in this area are expected to gain momentum with the recent free-of-charge offers of PHRs by Microsoft (HealthVault) and Google (Google Health).[21, 22] Both systems provide storage space for personal health information and services that allow data to be imported from many sources (e.g., hospitals, pharmacies, health devices) or shared with authorized parties. Data analysis and decision support can be obtained from other institutions upon user permission and through public application programming interfaces (APIs).

DiG guidelines are supposed to contain exclusively data interpretation steps for patients, not specific clinical workflow information. This approach encourages the creation of guidelines that are distributed documents whose parts are hosted by different institutions.

The ability to create guidelines in the form of distributed documents with the aid of semantic Web technologies is a key difference between DiG and other systems and brings two major advantages to patient-centric decision support:

- It enables prompt reuse of knowledge generated by specialized institutions for the benefit of the general population.
- It enables rapid dissemination of updates.

Sharing Knowledge with DiG: Case Scenario

DiG belongs to the broad domain of decision support frameworks for healthcare, a field with a long history of tools for assisting doctors and nurses. DiG's novelty derives largely from its focus on patients as opposed to health professionals and from the way expert knowledge is structured in it. The following hypothetical scenario illustrates the functionalities of the proposed system.

Institution A is a recognized center of excellence in matters of occupational diseases. In order to increase awareness of risks and preventive actions related to complaints of arm, neck, or shoulder (CANS), the institution decides to create a DiG guideline providing personalized risk assessment and recommendations for the population. The guideline contains a series of computer-executable rules involving technical concepts that need to be unambiguously understood for correct use. In order to prevent confusion in concepts such as obesity, the guideline authors refer to another guideline, maintained by Institution B, that defines rules for establishing obesity status based on body mass index (BMI). Both guidelines are freely available online and hosted in distinct servers belonging to the authoring institutions. For this reason, the authors of the CANS guideline do not have to repeat the work done by Institution B.

John, a worker concerned about his CANS risk status, decides to follow the procedure described in the CANS guideline. Because the guideline is machine executable, John only needs a computer application that interacts with his personal health record to access the CANS guideline Web site and request an automatic assessment. John's PHR system will download the

guideline, parse it, and apply its rules to the contents of his PHR. As some rules depend on the evaluation of John's obesity status, the BMI guideline from Institution B will be automatically located and used for this purpose. After successful completion of the process, John will be given a personalized evaluation and recommendations.

As illustrated in the example, a central element in the DiG approach is the concept of executable knowledge, which denotes any document containing facts and relationships in a format that allows logical consequences to be uncovered automatically by machines. To facilitate the creation and extension of domain knowledge for patient application, a model for encoding expert knowledge in an executable format is proposed. The model is essentially composed of a flowchart in which the steps include logical rules that operate on an underlying health evaluation model expressed as a directed graph. The resulting encoding of expert knowledge is structured in a document referred to as a *consumer guideline*. The document may exist in a distributed format on the Internet and be maintained by different organizations with distinct specializations. Knowledge integration occurs at execution time using the most up-to-date information available.

The use of semantic web technologies is essential for the aforementioned integration. Ontologies provide the vocabulary and meaning for consumer guidelines. Each concept in a guideline can be linked to a specific ontology and unambiguously identified by a uniform resource identifier (URI). Such an approach allows for the rigorous use of concepts in the guidelines.

System Design

DiG comprises a set of software modules that acts as a knowledge broker between experts and consumers as depicted in Figure 2. Domain experts produce machine-executable knowledge in the format of consumer guidelines that may exist as distributed documents on the Internet. Hyperlinks tie the pieces of the document together. The execution of the guidelines relies on PHR data and further software components (represented as "code" in the figure) to apply abstract knowledge to concrete cases. Consumers interact with framework-based applications through a Web browser to obtain personalized health insights (e.g., risk assessments) extracted from this knowledge. The framework coordinates each step of this process.

DiG is organized in a modular architecture composed of three layers as depicted in Figure 3. A clear separation of responsibilities between layers is maintained as outlined in the remainder of this section.

1. Value-Added Services Layer

This layer contains modules that offer clean interfaces to end-user applications by encapsulating all the details of lower layers. Examples of services provided by these modules include risk assessment for a specific disease or transformation of raw sensor data into language accessible to the user.

2. Expert Knowledge Layer

Placed immediately below the value-added services, this layer is the core of the proposed framework and is responsible for the execution of consumer guidelines. The engine in charge of parsing consumer guidelines and applying their codified knowledge to health data is referred to as Health-Guru. Applications, through value-added services, may invoke Health-Guru by providing a pair of values: a patient identification and a consumer guideline reference. The former enables the engine to access personal health data, and the latter represents the knowledge to be applied to such data.

Certain guidelines may optionally initiate a dialogue procedure in order to obtain facts directly from the user. This dialogue assumes the format of a set of questions that the user must answer before the evaluation proceeds. Multiple such dialogues are possible in a single guideline. The end result of the Health-Guru process is a set of recommendations for the user.

Consumer Guideline Model. Consumer guidelines are structured documents that allow expert knowledge to be easily codified into machine-executable format. The aim of a guideline is to control the process of collecting and interpreting facts that are relevant to a specific health assessment. A fact is defined as a triple being composed of subject, predicate, and object. The statement *Joe has diabetes* or (Joe, has, diabetes) thus constitutes a fact where *Joe* is the subject, *has* is the predicate, and *diabetes* is the object.

Consumer guidelines codify expert knowledge declaratively (i.e., as facts) to facilitate human readability. As a result, domain experts can more easily translate their knowledge into machine-executable documents. The relation between facts is stated in declarative rules of the following format:

If conditions then consequences or conditions → consequences

Here, `conditions` and `consequences` are collections (conjunction) of facts. Accordingly, the following rule establishes the relation of Joe's weight and height with his obesity condition:

Joe hasWeightInKg 80, Joe hasHeightInCm 155 → Joe is Obese

To be reusable by different patients, consumer guidelines cannot hardwire all facts as in the previous example. Instead, they should provide *fact templates* that can be instantiated to particular individuals at execution time. Templates are created with the aid of *virtual variables*. These coding elements allow the construction of generic statements that are applicable to multiple patients.

The result of a guideline step is a set of facts about the patient. As fact creation may depend on preexisting facts, an ordering in this process is required. The method is akin to interviewing an expert on a particular topic: the interview is dynamically organized based on what is known, and steps are skipped if not applicable. This ordering process is emulated by the consumer guideline model through the use of constructs representing evaluation steps. Eight step constructs are available to order and organize the fact inference process:

1. **Start**: marks the beginning of a consumer guideline algorithm.
2. **Abstraction**: extracts facts from PHRs and/or from previous steps of the guideline and derives new facts based on a number of codified rules.
3. **Dialogue**: interviews the user to extract facts that are not present in the PHR.
4. **Branch**: defines tests to select the next guideline step to be executed.
5. **Nest**: enables guidelines to embed other guidelines, thus facilitating reuse of expert knowledge.
6. **Load**: allows reusing a set of facts stored in a database or PHR.
7. **Advice**: used for returning information to the user (e.g., multimedia files, hyperlinks, plain text).
8. **End**: marks the end of a consumer guideline algorithm.

Steps in a guideline are organized as a directed graph where links between nodes (steps) correspond to their relation. Each step (with exception of *End*) uses a `hasNextStep` property to indicate the next step to be executed. Branches indicate multiple steps using the `hasConditionalNextStep` property. A guideline algorithm encoded using these concepts can be visualized as a flowchart. Declarative rules are embedded through the `hasRules` property.

All elements in the guideline, including instances of steps, are identifiable using a URI, which allows the composition of distributed documents over the Internet. Any element can be referred to through a hyperlink to its unique URI.

A graphical representation of the consumer guideline ontology is shown in Figure 4.

Health-Guru. A consumer guideline is parsed and executed by the Health-Guru module. During execution, the module progressively accumulates a collection of facts about the patient's condition. New facts are added to the collection based on PHR data, codified expert knowledge, logical inference, or computation. The consumer guideline indicates which facts to add and the order in which the facts are created. The result of the process is referred to as a health evaluation model and constitutes a snapshot of the relevant knowledge required to answer specific questions about the patient's health. The model can be represented as a directed graph where nodes are concepts (e.g., patient, overweight) and links establish a relation between these concepts (e.g., patient is overweight).

3. Persistence Layer

Data are retrieved from PHRs through methods provided by the persistence layer, whose arguments and results conform to a predefined patient data model (see Figure 5). Upper layers manipulate personal health data only through these objects and are, for that reason, shielded from the underlying database schema. The patient data model is a description of a person from the point of view of his or her health information. Name, age, weight, blood type, and other concepts are used in this description, which provides a basic common vocabulary for personal health information in consumer guidelines. As long as guideline authors describe patient data in terms of this model and with the aid of virtual variables, the persistence layer will be able to map these concepts into a PHR system and vice versa.

Discussion

Authoring consumer guidelines can be more or less challenging depending on factors such as previous experience, expertise in the domain knowledge being modeled, and technical background.

In order to evaluate the expressivity of the proposed consumer guideline model and language, a strategy inspired by the work of the GLIF team was used.[23] A test guideline containing knowledge required to perform a risk assessment of developing coronary heart disease (Framingham method) was codified by three different people (including the authors of this article).[24] The Framingham method computes a risk score by considering a person's age, total cholesterol, HDL cholesterol, systolic blood pressure, smoking status, and the use of medication for blood pressure control.

After codifying the Framingham risk assessment method, encoders provided feedback on the expressivity of the guideline model and language. The encoders used different steps to implement the same Framingham guideline. One reason for this difference lies in the way guideline authors extract facts from users.

Although not enforced by the guideline model, the encoders considered it helpful to follow a three-tiered approach to design consumer guidelines in which the first tier contains concepts related to the guideline model itself (e.g., *Steps*), the second tier deals with domain concepts (e.g., *Disease*), and the third tier contains the guidelines created with concepts from the first two tiers. Feedback from the encoders suggests that the framework should enforce a clear demarcation between tiers so that guidelines can be more portable.

The use of Protégé as a modeling tool was recommended by all three encoders, who agreed that a graphical user interface (GUI) customized for creating consumer guidelines would greatly facilitate the authoring process.[25]

Overall, encoders considered the model fairly intuitive, although their opinions were possibly influenced by the fact that they had good prior knowledge of ontology-based data modeling.

Conclusion

DiG is a framework intended to enable PHR applications to extend decision support to the general population. Once encoded, guidelines are easily shared, extended, and modified. These guidelines can exist as distributed documents on the Internet and be executed by a processing engine to provide an individual with personalized assessments against various health risks based on the evidence data stored in a PHR.

A critical area for future work is the representation of expert-domain concepts in a standardized way. This step is an essential condition for creating a large body of medical knowledge that can be easily extended. Another important area is the creation of graphical tools to ease the creation and sharing of consumer guidelines.

The adoption of a distributed-document model, despite its benefits for creating and updating guidelines, can also pose shortcomings. Strategies must be developed to ensure that the failure of single guideline components does not easily make entire classes of guidelines unavailable or erroneous.

In spite of the work ahead, the present work represents a first step in the direction of empowering the general population through access to accurate, easy-to-use health information.

Andre Barroso, PhD, is a senior research scientist at Philips Research Laboratories in Eindhoven, The Netherlands.

Edwin Yaqub, MSc, is a scientific assistant at the Technical University of Dortmund, Germany.

Acknowledgments

Notes

1. Tang, P. C., et al. "Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption." *Journal of the American Medical Informatics Association* 13 (2006): 121–26.

2. Markle Foundation. *The Personal Health Working Group Final Report*. 2003. Available at http://www.connectingforhealth.org/resources/final_phwg_report1.pdf (accessed May 2010).

3. National Institutes of Health. *Health Information*. Available at http://health.nih.gov/ (accessed March 2010).

4. NHS Health Development Agency. *Effectiveness of Public Health Campaigns*. June 2004. Available at http://www.nice.org.uk/aboutnice/whoweare/aboutthehda/hdapublications/effectiveness_of_public_health_campaigns_choosing_health_briefing.jsp (accessed May 2010).

5. U.S. National Library of Medicine. PubMed Home. Available at http://www.ncbi.nlm.nih.gov/pubmed/ (accessed March 2010).

6. Eccles, M., and J. Grimshaw (Editors). *Clinical Guidelines from Conception to Use*. Abingdon, UK: Radcliffe, 2000.

7. Mann, Tom. *Clinical Guidelines: Using Clinical Guidelines to Improve Patient Care within the NHS*. , UK Department of Health,1996.

8. National Guideline Clearinghouse. *About NGC*. Available at http://www.guideline.gov/about/about.aspx (accessed March 2010).

9. National Institute for Health and Clinical Excellence. Available at http://www.nice.org.uk/ (accessed March 2010).

10. Hunt, D. L., et al. "Effects of Computer-Based Clinical Decision Support Systems on Physician Performance and Patient Outcomes: A Systematic Review." *JAMA: The Journal of the American Medical Association* 280 (1998): 1339–46.

11. McDonnell Norms Group. "Enhancing the Use of Clinical Guidelines: A Social Norms Perspective." *Journal of the American College of Surgeons* 202 (2006): 826–36.

12. Wells, Susan, et al. "Patterns of Adoption and Use of a Web-Based Decision Support System in Primary Care (PREDICT CVD 7)." *Health Care and Informatics Review Online* (October 2007). Available at http://www.hinz.org.nz/journal/2007/10/Patterns-of-adoption-and-use-of-a-web-based-decision-support-system-in-primary-care-PREDICT-CVD-7/969#2 (accessed March 2010).

13. Majumdar, S. R., F. A. McAlister, and C. D. Furberg. "From Knowledge to Practice in Chronic Cardiovascular Disease: A Long and Winding Road." *Journal of the American College of Cardiology* 43 (2004): 1738–42.

14. Cranney, M., et al. "Why Do GPs Not Implement Evidence-Based Guidelines? A Descriptive Study." *Family Practice* 18 (2001): 359–63.

15. Langley, C., et al. "Use of Guidelines in Primary Care—Practitioners' Perspectives." *Family Practice* 15 (1998): 105–11.

16. Musen, M. A., et al. "EON: A Component-Based Approach to Automation of Protocol-Directed Therapy." *Journal of the American Medical Informatics Association* 3 (1996): 367–88.

17. Ohno-Machado, L., et al. "The GuideLine Interchange Format: A Model for Representing Guidelines." *Journal of the American Medical Informatics Association* 5 (1998): 357–72.

18. de Clercq, P. A., et al. "Design and Implementation of a Framework to Support the Development of Clinical Guidelines." *International Journal of Medical Informatics* 64 (2001): 285–318.
19. *SAGE Project Overview*. Available at http://sage.wherever.org/projectoverview/projectoverview.html (accessed March 2010).
20. OpenClinical. Methods and Tools for Representing Computerised Clinical Guidelines. http://www.openclinical.org/gmmsummaries.html/ (accessed May 2010).
21. *Microsoft HealthVault*. Available at http://www.healthvault.com/ (accessed March 2010).
22. *About Google Health*. Available at http://www.google.com/intl/en-US/health/about/index.html (accessed March 2010).
23. Ohno-Machado, L., et al. "The GuideLine Interchange Format: A Model for Representing Guidelines."
24. National Heart, Lung and Blood Study and Boston University. *Framingham Heart Study*. Available at http://www.framinghamheartstudy.org/ (accessed March 2010).
25. Gennari, J. H., et al. "The Evolution of Protégé: An Environment for Knowledge-Based Systems Development." *International Journal of Human-Computer Studies* 58 (2003): 89–123.
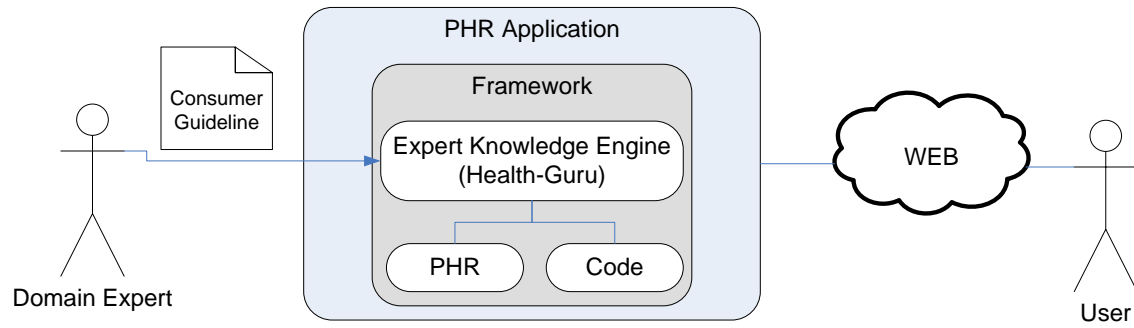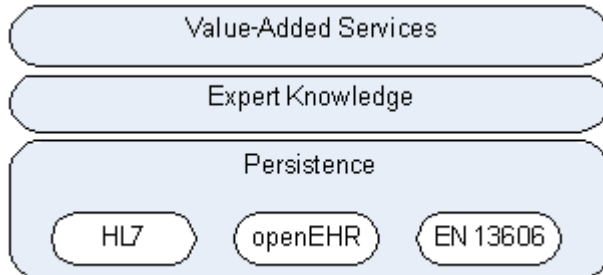
Figure 1

Clinical Decision Support System Diagram
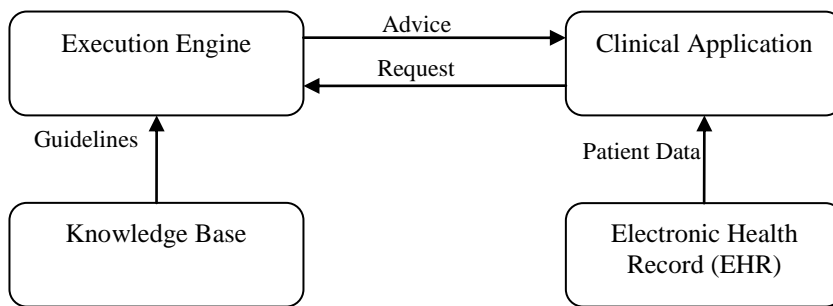
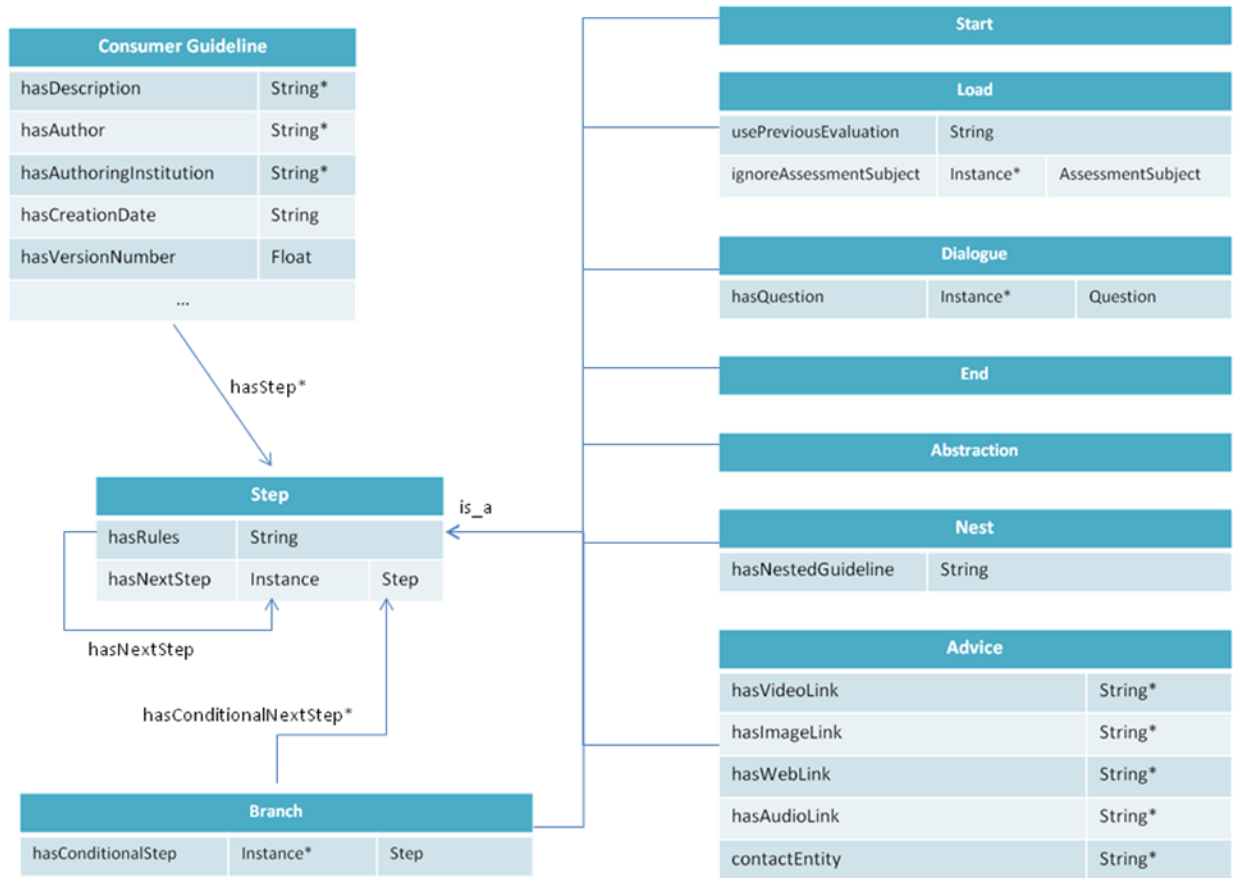**Figure 2**

DiG-based Personal Health Record (PHR) System

Figure 3

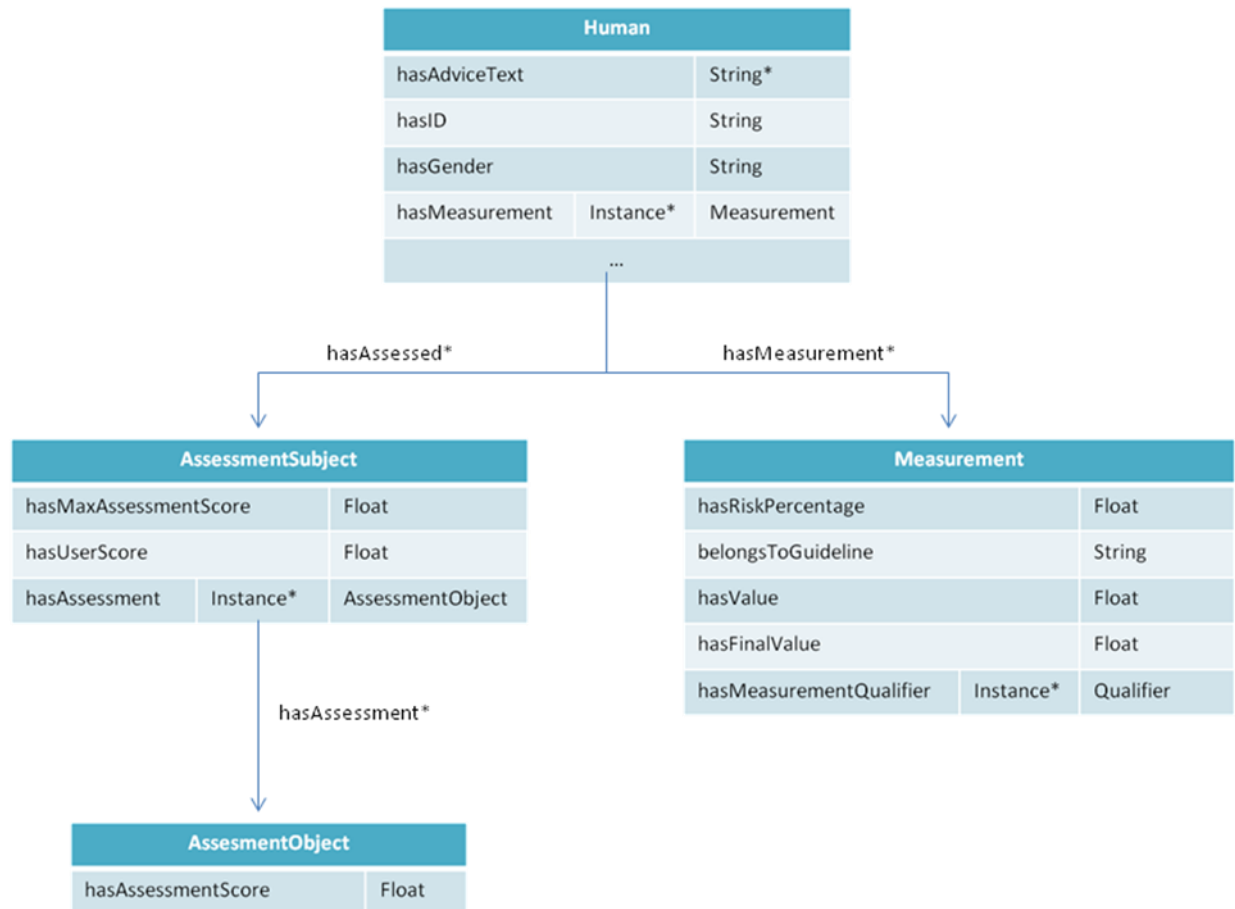Framework Stack

Figure 4

Consumer Guideline Ontology

Note: Guidelines are organized as a sequence of steps. Each step (with exception of *End*) uses a `hasNextStep` property to indicate the next step to be executed. Branches indicate multiple steps using the `hasConditionalNextStep` property. Declarative rules are embedded through the `hasRules` property.

Figure 5

Patient Data Model

Human		
hasAdviceText		String*
hasID		String
hasGender		String
hasMeasurement	Instance*	Measurement
...		

hasAssessed*   hasMeasurement*

AssessmentSubject		
hasMaxAssessmentScore		Float
hasUserScore		Float
hasAssessment	Instance*	AssessmentObject

Measurement		
hasRiskPercentage		Float
belongsToGuideline		String
hasValue		Float
hasFinalValue		Float
hasMeasurementQualifier	Instance*	Qualifier

hasAssessment*

AssesmentObject	
hasAssessmentScore	Float

Note: *Human* represents the person for whom the health assessment is being performed. *AssessmentSubject* is an aspect of *Human* health that is assessed by the consumer guideline (e.g., Does *Human* experience neck pain? What kind of smoker is *Human*?). *AssessmentObject* represents the possible outcomes of a certain *AssessmentSubject* (e.g., *Human* can be a heavy smoker, light smoker, or nonsmoker). An *AssessmentObject* usually carries a numerical score that represents its impact in the assessment. *Measurement* represents measurable attributes of a *Human* instance (e.g., height, weight, body mass index, risk of coronary heart disease).