eHealth Interoperability with Web-based Medical Terminology Services - A Study of Service Requirements and Maturity

Neil Barrett and Jens H. Weber-Jahnke
University of Victoria, Department of Computer Science, Victoria, British Columbia, Canada
Email: {nbarrett, jens} @uvic.ca.

Abstract—Interoperability of health information systems requires common clinical terminologies and services that make those terminologies available on a shared infrastructure, i.e., Web-based terminology servers. Terminology servers share characteristics with ontology servers, but also provide further services, e.g., extended linguistic functionality. There is increasing interest in research and development of Web-based clinical terminology servers. While several systems have been developed, few studies exist comparing their functionality and maturity. This paper identifies a list of service requirements based on a systematic literature review and uses these requirements to perform an assessment of current maturity of terminology servers. It further identifies which terminology services are sufficiently mature to be standardized and shared in large interoperable ehealth system infrastructures. This paper also exposes concerns and research directions related to shared terminology servers such as Web-based servers, e.g., in areas of privacy and security.

Index Terms—eHealth, terminology server, interoperability, SNOMED-CT

I. INTRODUCTION

Being a knowledge-intensive domain, health care relies on effective methods and infrastructure for communicating and sharing information among the various agents involved in the care process. For example, family physicians often share health information with specialists when referring patients. eHealth systems used by these different agents need to be interoperable in order to enable such information sharing. These systems are critical for modern health care processes [1], [2].

Various components contribute to health information system interoperability. Semantic interoperability is enabled by the use of shared clinical terminologies. Generally, a terminology can be defined as a “body of terms used with a particular technical application in a subject of study, theory, profession, etc.” [3]. A clinical terminology is a terminology used in the medical domain. Similar to the notion of semantic ontologies, terminologies may contain semantic relationships among terms. We refer to such terminologies as semantic terminologies in order to differentiate them from mere collections of terms.

Clinical terminologies can be accessed through software services. (Software services provide function to interested software client components.) While it is possible to implement and maintain distinct clinical terminology services for each health information system separately, sharing terminology makes good economic sense, because a largest portion of the cost of ownership of a terminology server is spent on maintaining and updating the knowledge base of such a system [3]. Sharing this cost among a large number of clients provides economy of scale and also further promotes interoperability by reducing the potential for inconsistencies among different versions of knowledge bases installed for different eHealth components. Although terminology services could be shared within an organization, greater impact results from sharing these as Web-services (Fig. 1). Making shared terminology services available requires an agreed upon set of services.

There is increasing interest in research and development of Web-based clinical terminology servers. While several systems have been developed, few studies exist comparing their functionality and maturity. This paper identifies a list of service requirements based on a systematic literature review and uses these requirements to perform an assessment of current terminology server maturity. It further identifies which terminology services are sufficiently mature to be standardized and shared in large interoperable eHealth system infrastructures. This paper also exposes concerns and research directions related to shared terminology servers such as Web-based servers, e.g., in areas of privacy and security.

The rest of this paper is structured as follows. The next section provides a more detailed introduction to semantic terminologies and their Web-based applications. Section 3 identifies service requirements for semantic terminology servers in the eHealth domain based on the results of a systematic literature review. Section 4 introduces the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) as an example for a specific semantic terminology in the eHealth domain. The SNOMED CT documentation includes a technical implementation guide (TIG) with information pertaining to the implementation of terminology services. Section 5 aligns the services...
described in the TIG with the service requirements extracted in our systematic literature review in order to identify potential gaps. Section 6 surveys concrete SNOMED CT terminology server implementations and aligns their capabilities with the extracted service requirements in order to identify potential gaps and indicate directions for research and development.

We provide a discussion of our study in Section 7 before closing with concluding remarks in the final section.

II. SEMANTIC TERMINOLOGIES AND THE WEB

A terminology is a set of terms with a declared meaning relevant to a particular domain. This base definition can be extended to include alternative term descriptions and synonymous terms. For example, a Web terminology might include the term *cascading style sheet* and its synonym CSS (also an acronym). Other terminology extensions are possible, e.g., multilingual descriptions.

Semantic terminologies include computer-interpretable semantics relating terms and are based on ontological models. An ontology is described as an explicit specification of belief representing agreed upon domain semantics [4]. An ontology facilitates information sharing between computing systems and semantic computation.

The Semantic Web incorporates ontologies into the World Wide Web. Ontological models on the Web enhance information searches, information presentation and electronic commerce [5]. They support data exchange and storage, reasoning and navigation [6]. Since semantic terminologies are based on ontological models, these terminologies can fully participate in a Semantic Web.

Introducing a clinical semantic terminology to the Semantic Web will enable the Semantic Web to be “health smart”. A study by the Pew Internet & American Life Project [7] reports that a significant number of adults search the Web for health information. A “health smart” Web may become important for consumer health and well being. While current clinical semantic terminologies have been developed specifically to be used by medical professionals, emerging terminologies to be used in consumer health applications will include terms that are more broadly accessible to laypersons.

III. SERVICE REQUIREMENTS

A terminology server is “a resource that delivers a range of terminological services” [8]. It provides terminology services that allow interested software access to the terminology. Service requirements describe and scope this interaction between a server and client components and are thus important for implementing an interoperable eHealth infrastructure. Component-based reuse of software becomes practical on a broad scale only if there is a shared, mature understanding of the functionality to be performed by such a component [9]. Software engineers have developed such a shared understanding for many different kinds of software components over the last several decades, e.g., database management systems, Web servers, email servers, versioning systems etc. Joint declarations of service requirements for such components (sometimes referred to as “manifestos”) have significantly contributed to focusing development and implementation efforts, e.g., [10]. This paper seeks to contribute to the development of a more concrete understanding of the requirements for terminology servers in order to assess and further their maturity for component-based reuse.

Many scholarly publications mention terminology server requirements. In order to collect a thorough list of these published requirements, we conducted a systematic literature review (February 2009) and extracted terminology server requirements from the body of identified publications. The review’s question was: what requirements are relevant to a terminology server’s functionality?

IEEE explore, ACM’s Guide, PubMed and EBSCO (Computer Science Index, CINAHL with Full Text) were searched for English publications using the three search phrases: “terminology server” and (architecture or design); “ontology server” and (architecture or design); “terminology server”. The search was broadened starting from the first search phrase ending in the third because few results were returned.

A total of 191 publications were found including duplicates; 13 were kept for review [4], [8], [11]-[21]. Publications were kept (inclusion criteria) if they discussed requirements, software components, design, architecture or application programmer interfaces. They could be qualitative, quantitative, observational, feasibility and characterization studies. A publication was evaluated by first reading the title, then the abstract and finally the full paper; stopping when it could be ascertained that the pub-
liciation did not match the review's objectives. Seven publications discussed terminology servers and five discussed ontology servers. Publication references were examined and appropriate papers fetched (limited to papers between the year 2000 and 2009). One additional paper was found (Hogarth, Gertz, and Gorin [22]) in this reference analysis step.

Based on the service requirement categorization provided by Ahmad and Colomb [4] and Bechhofer, Goble, Rector, Solomon and Nowlan [8], a terminology server should include services for terminology management (e.g. addition, deletion, deprecation, verification, security) and access (e.g. querying, knowledge generation, natural language bridging). That is, **management services** are the services required to maintain correct functioning of the terminology server and **access services** are the services used for semantic computation. This new dichotomy is employed to categorize service requirements.

The following two sections synthesize the service requirements extracted and inferred (main reference provided) from the methodological literature review's publications.

### A. Management

- **Addition (mass)** [4]: Addition of concepts, relationships or other artifacts to the terminology. This could also include mass addition or importing from a source such as a text file.
- **Deletion (mass)** [4]: Deletion of concepts, relationships or other artifacts from the terminology. This could also include mass deletion given some input such as a text file.
- **Deprecation and replacement** [11]: Terminology artifacts may be used by older systems or may exist in stored data. Functionality is required to handle transitioning from older to updated terminology or handle legacy systems and their terminology versions.
- **Development (collaborative)** [4]: Although many terminology servers may focus on use, terminology servers should also account for terminology development. Development is the front-edge of a terminology life-cycle whereas deprecation and replacement are the back-edge.
- **Verification** [4]: Terminology servers should incorporate automated mechanisms for terminology verification, or at least facilitate manual verification. This ensures the terminologies' consistency and soundness which are crucial for systems reliant on the terminology server.
- **Security and privacy** [4]: Terminology servers may contain sensitive information, may require restricted access, may need to guarantee data integrity and may need to guarantee a certain level of availability. Security and privacy considerations mitigate dangers to and instability of the terminology server.
- **Extensibility** [4]: Terminology servers should provide mechanisms for addition, removal and deprecation of software modules increasing the server's extensibility.
- **Usage and resource monitoring** [12]: Terminology servers should track their usage. This can lead to system optimization or detection of intruders.
- **Resource allocation, distribution and scheduling** [12]: Terminology servers should manage their resources in order to fulfill requirements (e.g. increased speed or reduced storage) and provide mechanisms to tune allocation, distribution and scheduling of their resources.
- **Inventory** [15]: Provides basic information about the terminology server such as the available versions.

### B. Access

- **Querying** [17]: Querying begins with basic requests for information and evolves to complex requests that border on knowledge generation and natural language bridging (see below). The following is a non-exhaustive list of more detailed querying service requirements synthesized by Chute, Elkin, Sherertz and Tuttle [17]; the last two are considered optional by the authors:
  - Enumeration of lexicon concepts and their associated information.
  - A means of retrieving a unique external identifier for each concept.
  - A means to retrieve attribute values which support translation to other coding schemes.
  - Enumeration of attribute types and relationships supported within a lexicon.
  - Enumeration of concepts which participate in specified relationships (e.g. hierarchical) with respect to a concept.
  - Enumeration of attributes and relationships for a given concept.
  - Enumeration of concepts corresponding to a specified attribute value.
  - Enumeration of concepts that satisfy multiple relationships and attribute value combinations.
  - Support partial pattern matching or generic selection on matching criteria.
  - Support traversal of relationships within a lexicon.
- **Knowledge generation or inference** [14]: Generating terminological information that does not otherwise exist.
- **Natural language bridging** [17]: Functionality that bridges the divide between natural language and the terminology. (e.g. translating natural language to terminology, multi-language). The following is a non-exhaustive list of more detailed natural language bridging service requirements presented in Chute, Elkin, Sherertz and Tuttle [17]:
  - Word normalization, such as simplifying variants *coughs* and *coughed* to *cough*.
  - Word completion, such as pn* implying pneumococcal, pneumocystis.
  - Spelling correction.
IV. SNOMED CT

In the rest of this paper, we will focus on a specific semantic clinical terminology, namely Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT). SNOMED CT "is a comprehensive clinical terminology that provides clinical content and expressivity for clinical documentation and reporting. It can be used to code, retrieve, and analyze clinical data" [23, page 8].

SNOMED CT contains a set of concepts, terms and relationships. Multiple terms from multiple languages may be associated to each concept. The concepts are also associated to natural language descriptions. Concepts are related, with the primary relationship being hierarchical (Fig. 2). The hierarchy provides concept granularity from the general to the specific. For example, the concept procedure on lymph node is more generic than the concept procedure on femur.

Concepts may have attributes that are themselves concepts. Attributes participate in modeling of concept definitions by restricting a concept's semantics. For example, the concept fracture of femur has an attribute finding site that restricts the concept's location to a femur.

A concept's attributes are subject to hierarchical restrictions. Given two concepts where one is more specific than the other, the specific concept inherits the generic concept's attribute restrictions. For example, a more specific concept than fracture of femur inherits the location restriction of femur. Each concept may further restrict its inherited attributes.

A SNOMED CT expression is a collection of concepts that represent an instance of a clinical idea. It can capture semantics beyond those explicitly defined by SNOMED CT. Consequently, SNOMED CT is capable of expressing more than its individual concept semantics.

As a semantic terminology, SNOMED CT allows for concept normalization and post-coordination. Post-coordination is a process for generating complex expressions such as those not explicitly defined by SNOMED CT. Normalization transforms an expression to a standard form. One use of normalization is expression comparison (e.g. equality or subsumption). Normalization and post-coordination enable semantic manipulation of clinical concepts.

A more complete description can be found in the SNOMED CT User Guide [23] and additional information can be found on the International Health Terminology Standards Development Organization (IHTSDO) website: www.ihtsdo.org. IHTSDO develops and promotes the use of SNOMED CT.

Canada Health Infoway (www.infoway-inforoute.ca) adopted SNOMED CT for pan-Canadian use in electronic health systems. This adoption has increased the value in understanding and communicating analyses of SNOMED CT research within Canada and other countries similarly positioned (see www.ihtsdo.org/members).

Canada Health Infoway is a not-for-profit organization that works with the Canadian provinces and territories, health care providers and technology solution providers to accelerate the Canadian use of electronic health records.2 In doing so, Canada Health Infoway expects to provide health care teams, potentially dispersed over great distances, the infrastructure they need to share information and collaborate to make better-informed decisions about patient diagnosis and treatment.

C. Scenario

Having described SNOMED CT, a clinical scenario is presented to illustrate its use. The scenario is based on Giamangelo and Fenton's publication [24] that identifies several current and future applications of SNOMED CT. Their findings are summarized into four categories:

- **Data collection.** SNOMED CT participates in data collection such as encoding a physician's written notes.
- **Computer interpretation of data to support health care practice.** SNOMED CT enables software to compute on health data. This could include clinical guidelines associated with a patient's record enabling health professionals to access the latest clinical guidelines relevant to patient treatment. It could also include intelligent alerts such as warning of potential adverse drug interactions.
- **Data communication.** Having two systems communicate health information, such as medical devices encoding data for communication to other interested devices.
- **Research.** Making data available for research through automated mechanisms. This could include measuring a clinical outcome across multiple data sets.

**Scenario:** Patricia is having difficulty swallowing so she makes an appointment to see her family physician, Dr. Don. During the appointment, Patricia and Dr. Don discuss Patricia's situation. Dr. Don refers Patricia to a throat specialist, Dr. Stephanie. After their appointment,
Dr. Don makes written notes in Patricia's electronic file. The computer automatically reads and encodes the notes in SNOMED CT. The computer is uncertain about one encoding and presents two possibilities to Dr. Don who chooses the correct SNOMED CT concept. The computer automatically communicates SNOMED CT encoded medical data required by Dr. Stephanie, to her. Dr. Stephanie examines Patricia and decides on a treatment plan. When she prescribes a medication, the computer alerts her of a possible minor side effect. The side effect is not a concern and the medication is prescribed. Since Patricia has already agreed to participate in a research study, Dr. Stephanie's diagnosis and prescription data is made available to Roger, a researcher. After several weeks Patricia is able to swallow properly. She accesses her personal health record on the Web and indicates as such. The computer automatically reads and encodes the notes and indicates statistical significance.

V. SNOMED CT TECHNICAL IMPLEMENTATION GUIDE MATURITY

The SNOMED CT Technical Implementation Guide (TIG) “provides technical implementation guidance to assist in the effective development of SNOMED CT Enabled Applications” [25, page 10]. Consequently, the TIG should assist in the effective development of a SNOMED CT terminology server, including guidance on the development of a terminology server’s services for interoperability. To assess the guide’s maturity, the TIG was read and each subsection was associated (when possible) with one or more service requirements (see Table I and II). In other words, for each section the associated service requirements were selected (Fig. 3). The pages spanned by each section were totaled and also appear in Table I and II. The document’s appendixes were excluded from this process.

No additional service requirements were discovered while reading the TIG that had not been discussed in the literature review.

With respect to service requirements, there is more content related to access service requirements relative to management service requirements. Within the management service requirements, there is more content related to addition, deprecation and replacement as well as extensibility.

The TIG did not comment on collaborative development, privacy and security, usage and resource monitoring, and resource allocation, distribution and scheduling. Consideration should be given to these service requirements, particularly when a clinical terminology is shared broadly. For example, implementing a terminology server as Web-services would require some level of privacy and security, usage and resource monitoring, and resource allocation, distribution and scheduling.

VI. SNOMED CT TERMINOLOGY SERVER MATURITY

Although the term server conjures notions of a networked client/server model, a terminology server may not be separated from its front-end. Some SNOMED CT browsers expose their terminology server as an application programmer’s interface (API). Under the assumption that browsers expose most terminology services to end-users as features, browser features indicate the maturity of the underlying terminology server (particularly to those who possess an understanding of the underlying functionality required to support a user-level feature).

To investigate SNOMED CT terminology server maturity, each SNOMED CT browser feature discussed in a review by Rogers and Bodenreider [26] was matched (when possible) with one service requirement presented in Section III. In other words, for each feature the best associated service requirement was selected (Fig. 3). The browser features were compiled by Rogers and Bodenreider upon inspection of the following systems (first 16 as working software): CaTTS, CliniClue, CLIVE, Ed-Browse, FDB Sphinx, HealthTerm, LexPlorer, Mycroft (Apelon), NCI Terminology Browser, OntoBrowser, OpenKnoME, Protégé-OWL, SNOB, Snoflake, the UMLS Rich Release Format Browser and the Virginia Tech Browser; HealthTerm, HLl LExPlorer and AxSys Browser. No additional service requirements were discovered during matching that had not been discussed in the literature review.

Implementations focus on access services relative to management services (see Table I and II). The greatest focus is querying.

VII. DISCUSSION

This paper suggested a set of terminology service requirements divided into two categories: management and access. Given these service requirements, a preliminary assessment of SNOMED CT terminology service maturity was performed. No additional service requirements were discovered during the SNOMED CT TIG and server maturity assessment process.

Both the TIG and terminology server implementations focused on access rather than management. The most developed services are querying and natural language bridging. (This is likely a manifestation of a terminology server’s user base, that of users rather than creators. Furthermore, even creators need to query their creations rather immediately as a basic form of evaluation.) Given that querying and natural language bridging are the most mature, service standardization for the purpose of interoperability could begin with these services. The existing documentation and practical experience could form the foundation for standardizing these services. This notion of maturity implying fitness for standardization could also be applied to the remaining services, providing a strategy for addressing service standardization.

Documentation and implementations do not address
privacy and security service requirements. Assuming these services are shared as Web-services (shared broadly), this research exposes privacy and security concerns. For example, observing server queries and patients accessing a clinic for a prolonged period could allow a malicious individual to correlate clinical data to individuals with high certainty. In such situations, a person’s privacy is comprised and personal health information becomes known. With respect to security, an individual could manipulate critical health information by gaining access to a terminology server. The individual could alter terminological information such as changing pneumonia codes to fractures altering automated procedures and disrupting health care processes. Since the privacy and security of personal health information is considered an important quality factor of eHealth systems [27], these are notable concerns for future research.

VIII. LIMITATIONS

A limitation of this research method is its subjectivity since a best association between service requirements and document sections or browser features is subject to interpretation. This interpretation was performed by an individual with about one year SNOMED CT knowledge and experience, which was deemed sufficient for such a process. This methodology produced an overview of SNOMED CT terminology server maturity. The methodology is limited in that it combines the features of all surveyed browsers. This implies that the individual terminology server is less mature than the picture presented here. Although these methods are sufficient for a preliminary assessment of SNOMED CT terminology server maturity, more rigorous methods should be employed in the future.

The TIG contained additional information beyond information related to the services, such as database implementation (sections 6.3.3 to 6.3.5 and 6.10), communication (6.10.3), user interface (7.2.2 to 7.2.9) and storage format (7.3) guidance. These sections were not associated to the suggested services. Similarly, several features such those under the headings visualization, navigation and usability were not associated with the suggested services. These features were visual in nature rather than being dependent on the underlying terminology server.

Although the suggested terminology services are based on previous research, this previous research does not necessarily reflect the context and requirements of healthcare based terminology services. It may be beneficial to study terminology server use within healthcare to gain a domain specific context and requirements knowledge.

IX. CONCLUSION

eHealth interoperability hinges, among other things, on the adoption of common terminologies to encode health information. Implementing such clinical terminologies in large scale eHealth systems in an effective and economic way constitutes an engineering challenge that requires a sound understanding of the requirements involved. Our study has identified a list of service requirements on clinical terminology servers based on a systematic literature review. This list of requirements may be used as a starting point to develop a “manifesto” on clinical terminology servers, which would help to focus further research and development activities in this domain. We also showed how this list of service requirements can be used to assess the completeness and maturity of existing terminology server standards and products, using specifically the SNOMED CT terminology. Clearly, our understanding of service requirements on clinical terminology servers is still imprecise and far from our level of understanding of other types of software components, e.g., database management systems. Particularly the question on how to query terminologies must be studied in more detail. Ontology query languages such as SPARQL [28] may provide a starting point but these lack many of the...
linguistic features required. Another topic warranting further research is the security and privacy impact of shared terminology servers.

REFERENCES


Neil Barrett received a master's degree in computer science from Memorial University of Newfoundland (2007) and a bachelor degree from McGill University, Montreal (1998). He is pursuing a doctorate degree from the University of Victoria.

Between his bachelor and master's degrees, he worked in industry on Web applications. He also completed several contracts for various organizations, from options trading to flight simulator configuration.

Mr. Barrett currently holds an NSERC scholarship and a University of Victoria President’s scholarship. He is a Fellow of the School of Graduate Studies of Memorial University of Newfoundland.

Jens Weber-Jahnke has received a doctoral degree in computer science from the University of Paderborn (1999) and a master's degree in Software Engineering from the University of Dortmund (1994), both in Germany.
He is the Director of the Software Engineering Program at the University of Victoria, in Canada, British Columbia. He is also an Associate Professor in the University’s Department of Computer Science and an Adjunct Associate Professor in the School of Health Information Science.

In 2005/2006, he has been a Visiting Associate Professor in the Department of Family Practice at the University of British Columbia. He has been an Industry Research Fellow with the BC Innovation Council since 2001. From 1994 to 1999, he has been a Research Associate in the Department of Computer Science at the University of Paderborn.

He has published extensively on Software Engineering and Data Engineering aspects related to health care. He is currently writing a forthcoming textbook on Engineering Clinical Information Systems to be published by Springer.

Dr. Weber-Jahnke is a senior member of the IEEE Computer Society and a senior member of the ACM. He is licensed as a Professional Engineer with the Association of Engineers and Geoscientists in the Province of British Columbia, Canada.