Towards a Flexible Semantic Framework for Clinical Trial Eligibility using Topic Maps

David Damen
biomina - biomedical informatics research center
Antwerp
Antwerp University Hospital
Wilrijkstraat 10
B-2650 Edegem, Belgium
david.damen@uza.be

Tim Van den Bulcke
biomina - biomedical informatics research center
Antwerp
Antwerp University Hospital
Wilrijkstraat 10
B-2650 Edegem, Belgium
-
ADReM - Advanced Database Research and Modelling
University Antwerp
Middelheimlaan 1
B-2020 Antwerp, Belgium
tim.van.den.bulcke@uza.be

ABSTRACT

Clinical trials are conducted extensively in clinical research. Most often, the eligibility screening consists of manually reviewing individual (patient) health records to identify potential candidates. Automated selection can improve the selection accuracy, increase the number of selected patients, and reduce the cost of the selection process. A formal representation of clinical trial eligibility criteria is however required to facilitate such automated selection and also to enable efficient multi-centric screening, thereby dealing with heterogeneous data coming from disparate sources.

We present a novel framework to represent clinical trial eligibility criteria based on Topic Maps. The framework addresses the needs in multi-centric screening by allowing for flexible and decentralized creation, sharing and integration of clinical trial representations and eligibility concepts. This is achieved through a formal but locally extensible ontology, thereby enabling unified querying across multiple institutions.

Topic Maps is a semantic knowledge representation technology for describing data and information. An ontology was created that describes the building blocks of a clinical trial. This representation encapsulates institution-specific data sources, thereby providing a domain-centric view on clinical trial definitions. Additionally, an evaluation engine was developed to assess patient eligibility using the ontology. A proof-of-concept was implemented at the Antwerp University Hospital. We show how the use of Topic Maps can facilitate patient recruitment across different centers.

Categories and Subject Descriptors

J.3 [**Computer Applications**]: Life and Medical Sciences—*medical information systems*; I.2 [**Computing Methodologies**]: Artificial Intelligence—*Knowledge Representation Formalisms and Methods*

General Terms

Standardization, Design

Keywords

Topic Maps, ontology, semantic web, clinical trial, eligibility criteria, decision support system, electronic health record

1. INTRODUCTION

When conducting a clinical trial, suitable participants need to be recruited by identifying eligible candidates for the trial, i.e. meeting a predefined set of in- and exclusion criteria. Eligibility screening usually occurs manually by reviewing individual health records. However, up to 60% of eligible candidates may be missed using this approach [12, 18], thereby prolonging the time required to meet the specified accrual target. The optimal recruitment of candidates for participation in clinical trials is of major importance for screening centers and study sponsors, being both a major revenue source as well as offering the potential of improved patient treatment. Studies have shown that the automated selection of eligible candidates could improve the selection accuracy [3, 5], increase the number of selected participants [12, 18], and reduce the cost of the selection process [6]. Moreover, clinical trials are conducted in a highly international and multicentric context [17], requiring more uniform representation formats for clinical trials to enable both exchangeability of eligibility criteria and multicentric querying.

Automated selection of eligible participants is also of great importance to feasibility surveys which are conducted by study sponsors prior to the start of a trial to estimate the number of patients that could be recruited at a potential site. Automated systems can aid to increase response times and provide more accurate estimations to these surveys.

While previous studies have focused on automated selection of eligible patients for clinical trials, most of these were prior to the introduction of EHR (Electronic Health Record) systems and with limited focus on semantic representations [20]: OncoDoc, OncoLink, EligWriter, ASPIRE, caMatch, T-Helper, EON, AIDS2, OaSIS, PROforma, Asbru, GLIF, SAGE, GUIDE, and PRODIGY. LinkedCT [10] is an existing open semantic web data source that collects clinical trial data and focuses on discoverability of clinical trials, but not on active patient recruitment.

For efficient formulation of eligibility criteria and for portability across different institutions, it is crucial to be able to re-use previously defined criteria. Examples of systems that incorporate this functionality include EligWriter, ERGO and ASPIRE. We refer to Weng et al. [20] for a complete overview of formal clinical trial representation systems. While a number of representation formats have been proposed (such as ERGO [19], ASPIRE [15] and CRFQ [1]), there is currently no broadly accepted standard for a knowledge representation of eligibility criteria [20]. Not all representation formats include a formal evaluation engine for candidate eligibility using real-life EHR data. Only few representation formats allow for the creation of derived eligibility concepts. Current formats also lack capabilities for representing criteria in more complex and decentralized multi-centric environments.

In this paper, we present a novel framework for the formal representation of clinical trial eligibility using Topic Maps [8, 16], including an automated evaluation engine. The evaluation engine is capable of integrating with existing local EHR systems, while providing a single abstracted engine for all clinical trials. The proposed framework allows for decentralized creation and sharing of eligibility concepts. Furthermore, we show how the use of Topic Maps addresses the goal of sharing clinical trial protocols between multiple centers through a formalized, locally adaptable ontology, thereby enabling unified querying across multiple institutions.

In the next sections, we first briefly review different semantic technologies for knowledge representation. Next, a short general overview of Topic Maps is provided before describing the complete design and use of our clinical trial ontology and evaluation engine. Finally, we discuss some of the implications of the proposed framework and illustrate the use of the ontology.

1.1 Semantic Technologies

The main goal of creating a semantic representation of a knowledge domain is to remove ambiguity in the interpretation of the concepts and ideas of that domain. The clinical trial domain is no different. Our aim in creating a semantic representation of clinical trials is to formally represent the trial's eligibility criteria in a machine-readable format that can be shared between centers and adapted to a specific infrastructure without loss or corruption of information.

1.1.1 RDF

The most widely used standard for semantic descriptions is RDF (Resource Description Framework), a W3C (World Wide Web Consortium) standard for representing Web resources and information about them [13]. It uses a predicate-based paradigm to make statements, typically referred to as a *triplets*, about resources. The RDF standard also provides formal semantics for the representation, enabling knowledge inference. To facilitate sharing of semantic representations

of information from the same knowledge domain, efforts are made to use the same structure for these semantic representations. Such a common structure is called an ontology. OWL (Web Ontology Language) is such a knowledge representation language for authoring ontologies of RDF representations [11].

RDF representations are typically made accessible from *triplet stores* and can be queried using SPARQL (SPARQL Protocol and RDF Query Language), a standardized query language for traversing RDF graphs. Here, the ontology that structures a triplet store can be considered analogous to a database schema for relational databases. Even though RDF is a widely accepted standard, there are some limiting aspects. The focus of RDF on using *addressable web resources* requires workarounds when dealing with more ephemeral concepts like medical conditions. The formal semantics of RDF also impose a stricter formalism which can be perceived as overly rigid by non-knowledge engineers such as clinicians and study nurses.

Linked Data is set of design guidelines for semantic data repositories meant to improve the discoverability of data [2]. It focuses on the use of dereferenceable URIs to identify concepts and on standards (RDF) to represent those concepts.

1.1.2 Topic Maps

Topic Maps (ISO/IEC 13250:2003) is set of standards comprising of: a) a reference model and a data model; b) textual representation formats in XML (Extensible Markup Language) and plain text; c) a graphical representation form; d) a constraint language; and e) a query language. This paper uses the convention of capitalizing terms used in the context of Topic Maps as a technology, and lowercasing terms used in the context of the *implementation* of Topic Maps technology. Several Topic Maps engines (e.g. Ontopia, Maiana, Wandora) exist for different programming languages that implement some or all of the above standards. The use of Topic Maps entails a subject-centric view of data, information and knowledge, with a clear separation between identification, naming and linking of concepts. Standardized support for merging topic maps, and for the scoping and reification of individual topic map constructs, facilitates the creation of flexible and dynamic semantic resources.

The basics of Topic Maps consist of the combination of *Topics*, *Associations* and *Occurrences*. *Topics* represent subjects, which are any kind of abstract (a recipe, a medical condition) or concrete (a person, a drug) concept. *Associations* define relations between different Topics. Finally, *Occurrences* serve to provide further information about a Topic, both as values inside a topic map, e.g. an address, or as a link to an external information source, e.g. a link to an image. Figure 1 shows a small example topic map illustrating these concepts. A more complete description can be found in Pepper et al. [16] and Garshol et al. [8].

This representation framework provides us with more flexibility than RDF. The subject-centric view that underlies Topic Maps provides a natural fit to describe a knowledge domain, such as the domain of medicine, in contrast to RDF, which was initially conceived for the creation of metadata (i.e. *statements*) about addressable web resources. RDF representations can still have ambiguity as to the precise target of the URI (uniform resource identifier): a URI that represents the concept *diabetes* can e.g. refer to either the medical condition 'diabetes' or to a webpage that describes
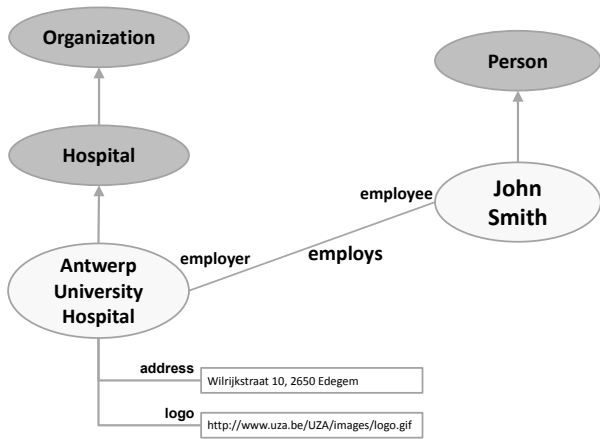
Figure 1: An example topic map.

diabetes.

Topic Maps on the other hand provides a framework to deal with such ambiguity. Furthermore, there are several additional properties of Topic Maps that are of particular interest for representing clinical trial eligibility:

IRI-based identification While Topics can be assigned one or more Names, these Names are not used for identification purposes. Topics are formally identified through three sets of subject indicators which are represented as IRIs (Internationalized Resource Identifier) thereby enabling flexible terminologies. For instance, when two topic maps would be merged, where each contains a Topic with *diabetes* as Name yet with no similar identifiers, then those Topics would be considered as different. As such, different centers can maintain their individual terminology, while still being able to link these terminologies with other centers in an easy and decentralized manner. Note that the use of IRIs as subject indicators for Topics and as locator for external Occurrences lets topic maps be 'good citizens' on the Semantic Web and in Linked Data spaces.

automatic merging The subject indicators of Topics form the basis of the standardized automatic merging algorithm for topic maps. Multiple topic maps can be merged seamlessly enabling a built-in support for knowledge fusion. This property is particularly interesting when merging topic maps from different institutions to enable multicentric querying for specific clinical trials.

scopes Scopes can be assigned to Names, Occurrences and Associations, and they indicate in what context those constructs are valid. A typical example of scope is language. The *Hospital* topic could have two names: 'hospital' scoped by the topic *en* (English) and 'Krankenhaus' scoped by the topic *ge* (German). A more advanced use of scopes would be scoping by time, i.e. scoping an association with a topic representing a time interval to indicate when the relation between certain topics was valid. Note that the latter is still an open research question in the domain of clinical trials [20]. This mechanism of scoping is not present in RDF.

reification Every name, association and occurrence in a topic map can become the subject of a new topic in that same topic map. This act is called **reification** and it allows for additional information to be added about topic map constructs. In our clinical trial ontology, this is used to link generic descriptions of clinical trials with hospital-specific information sources as in Figure 4 below.

These properties of Topic Maps provide us with a flexible framework from which we can create a formal, yet natural and easily adaptable machine-readable representation for clinical trial eligibility criteria for use across multiple centers.

1.1.3 Topic Maps ontology

In Topic Maps, an ontology is defined as the collection of topics to be used as types (topic types, association types, role types, occurrence types) in a topic map, together with a description of how topic instances are allowed to be associated with each other. The Topic Maps Constraint Language can be used to formally describe an ontology. Note that in Topic Maps, an ontology is itself again a topic map. Creating an ontology is typically an iterative process and often adapted to suit the specific needs and availabilities of the different stakeholders involved [4, 9].

2. METHODS

2.1 Overview

At the lowest level in our ontology, we define a *concept*, a basic primitive that represents specific information about a patient. Examples are: patient gender, lab test results, diagnostic codes. These concepts can be combined with *filters* and *comparators* into *cells*. A *cell* can be interpreted as a single logic statement representing a particular eligibility criterion of a clinical trial (see Figure 3). Multiple cells can be aggregated into *groups* through a combination of *logical operators*: AND, OR, NOT (see Figure 4). Groups can also contain other groups. *Clinical trial eligibility* is defined through a top-level *included* group that reflects if the subject is an eligible candidate for the complete *clinical trial*. It typically collects at least two other groups, one combining all inclusion criteria and one combining all exclusion criteria.

The rest of this section defines the structure for representing clinical trials in an ontology. The figures below illustrate the use of this ontology in a topic map for a fictional clinical trial, called CARDIOS. To refrain from overloading the example figures, we have omitted non-crucial information such as Roles where possible. A complete overview of the ontology can be found in Figure 7 in Appendix A.

2.2 Clinical trial ontology

A minimalist approach was taken in designing the ontology by only including definitions and constraints that are directly relevant to clinical trials, e.g. the *Institution* topic type requires only that a name is provided by its instances. While an *address* occurrence could be useful, it is omitted in this representation since it is not absolutely required. As mentioned, this ontology can easily be extended or linked with other ontologies to reflect such additional information.
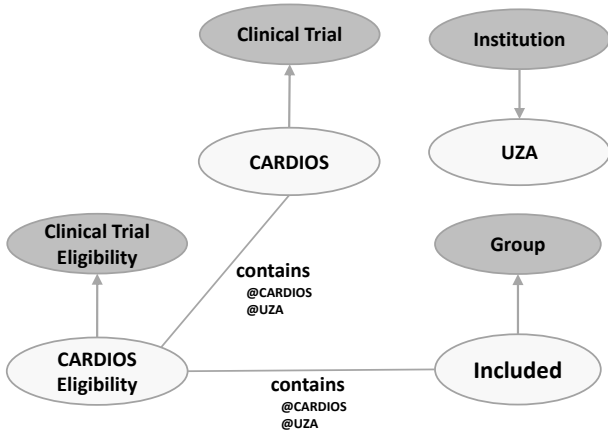
Figure 2: Example *Organization*, *Clinical Trial* and *Group* topics.

Figure 3: A *Group* composition example.

2.2.1 Institution, Clinical Trial and Group

The *Institution* and *Clinical Trial* topic types are used to respectively represent the centers at which clinical trials are conducted, and the clinical trials themselves. Instances of these types need only to provide a Name and are primarily used as scope on Associations.

The *Group* topic type is used to group a set of eligibility criteria that are combined in a logical rule, e.g. to group a set of demographic or hematological criteria into a new concept (e.g. 'heart problems' or 'diabetes'). Group instances only need to provide a Name for the group. For a specific topic map, a *Clinical Trial* instance will be linked to *Clinical Trial Eligibility* which in turn connects to an *Included* Group to serve as the entry point into the rest of the eligibility criteria representations through a *contains* association. Figure 2 shows example *Institution*, *Clinical Trial*, *Clinical Trial Eligibility* and *Group* topics.

2.2.2 Composing Groups

Next, the *Study Operand* topic type is defined as the supertype of the *Group*, *Cell* and *Logical Expression* topics. A *Group contains* a single *Study Operand*, which can be any of the above subtypes. If the *Study Operand* is a *Logical Expression*, then a *combines* association connects it with two other *Study Operands*. The *Logical Expression* topic type is itself a supertype of the following three logical operations: *AND*, *OR* and *NOT*. This mechanism enables the creation of a tree-structure representing the logical rule that combines the different eligibility criteria of a specific clinical trial. The leaves of this tree are always *Cells*. Figure 3 shows an example group, namely *CARDIOS Demographics*, that requires participants to the CARDIOS clinical trial to be both adult and female. As before, *contains* and *combines* associations will be scoped with the *Clinical Trial* and *Organization* instances for which they are valid.

2.2.3 Cells

The logical combination of *Groups*, *Cells* and *Logical Expressions* serves as the basis for the overall rule structure of a clinical trial. *Cells* also provide the link of that structure to hospital-specific information sources.
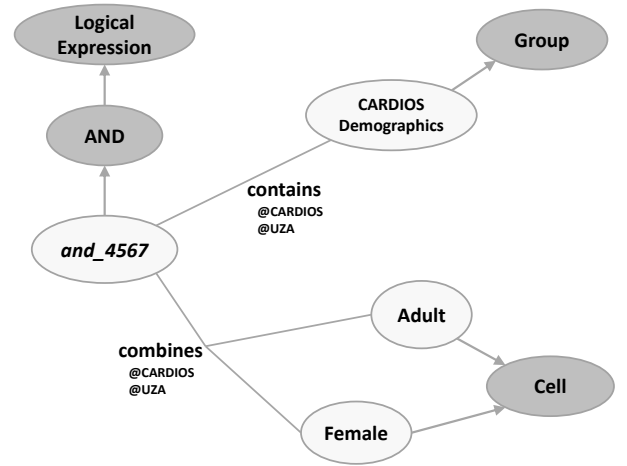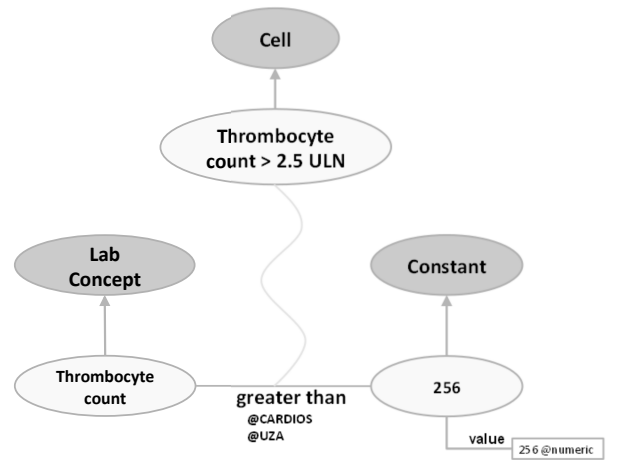

Figure 4: An example *Cell*.

A *Cell* instance reifies a *compares to* association between two *Operands*. The type *compares to* is a supertype for the following comparison operators: equals, greater than, greater than or equal, less than, less than or equal, ident. When creating a clinical trial topic map, those specific subtypes are used rather than the generic *compares to* association.

The example in Figure 4 illustrates how an abstract, structural *Cell*, representing a thrombocyte count well above normal, links the clinical trial representation to a concrete hospital infrastructure, i.e. that thrombocyte counts can be found as lab data, which are stored in a specific format and unit, should be higher than 256 to account for the way they are stored in this specific hospital.

The *Operand* topic type is a supertype for the *Constant*, *Filter* and *Concept* topic types. Instances of *Constant* have a name and a value, where the value is also scoped with *string* or *numeric* to indicate its value type. Instances of *Filter* reify a *filters* association, where *filters* is a supertype of the same comparison operators as *compares to* (with the exception of ident).
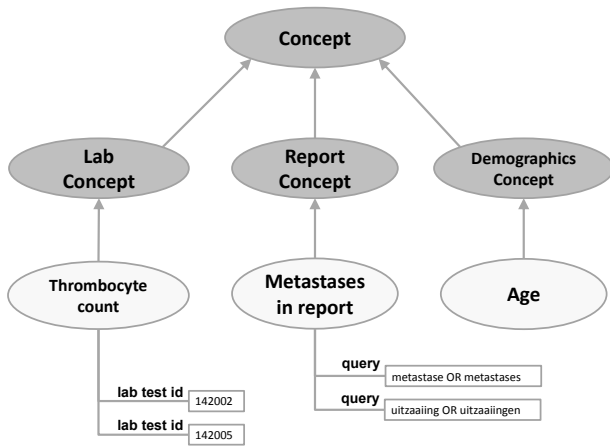
Figure 5: Example *Concepts*.

2.2.4 Concepts

Finally, there is the *Concept* topic type which is used to represent hospital-specific information sources. Subtypes of *Concept* distinguish between the various data sources that are available through the hospital IT infrastructure. Different subtypes require that instances provide different kinds of information in Occurrences. The following (yet incomplete) list of *Concept* subtypes illustrates the variety:

Lab Concept *Lab Concepts* represent lab values of patients such as thrombocyte counts or hematocrit values. Instances provide one or more *lab test id* occurrences that identify specific lab tests in the hospital. A *unit* occurrence can be provided for illustration purposes. A *selection* occurrence can used to indicate that only a subset of all lab values belonging to the *Lab Concept* for a specific patient should be taken into account, e.g. only the last value.

Diagnostic Code Concept *Diagnostic Code Concepts* represent diagnostic codes that have been assigned to patients. In Belgium, for instance, ICD-9-CM (International Classification of Diseases, Clinical Modification) codes linked to medical procedures are assigned.

Demographics Concept *Demographics Concepts* represent demographic information, such as age, gender and address. An additional *unit* occurrence can be provided.

Report Concept *Report Concepts* represent the occurrence of keywords in textual data sources such as patient discharge notes. Instances provide one or more *query* occurrences containing keywords that would need to appear in the textual data.

Medication Concept *Medication Concepts* represent the medication that has been administered to a patient in the medical center. Occurrences can contain *drug names* or *ATC (Anatomical Therapeutic Chemical Classification System) codes* of medication that patients receive through the hospital pharmacy.

2.3 Evaluation Engine

2.3.1 Open world/closed world assumption

The structure in clinical trial topic maps adhering to the clinical trial ontology also forms the basis for an evaluation engine to automatically identify suitable candidates among the patient population in a screening center. The expression of absent or negative information poses some specific challenges when formulating the eligibility criteria. Under a closed world assumption, if something is not known to be *true*, it is considered as *false*. For clinical trials, an eligibility criterion '*last HbA1c measurement > 100 nmol/L*' would therefore lead to *false* if there are no lab measurements for a particular patient. However, we expect the criterion to evaluate to an inconclusive state rather than true or false since the statement can still be either true or false (depending on the outcome of the lab test).

In an open-world assumption, the truth value of an eligibility criterion does not depend on if it is known or not by a specific observer. This implies that something is considered *unknown* unless it is explicitly stated as either *true* or *false*. This open-world assumption matches better with the domain of clinical trials and is therefore adhered to here.

2.3.2 Ternary Kleene logic

Our evaluation engine processes a *Logical Expression* using a ternary logic based on Kleene logic [7] that uses the following truth states: *true*, *false* and *unknown*. In this logic, *unknown* can be considered as either *true* or *false*. Logic operations that involve an *unknown* value and which are unambiguously either *true* or *false*, are considered *true* or *false* respectively in Kleene logic. E.g. '*true or unknown*' would resolve to '*true*' and '*true and unknown*' to '*unknown*'.

2.3.3 Rule Evaluation

The input for the evaluation engine is a clinical trial topic map, an *Institution* topic as scope, and a list of patient identifiers. The engine then performs a postorder traversal of the criteria tree.

When a *Cell* instance is encountered, the engine calls a (center-specific) service for the particular *Concept* involved, thereby providing the service with the patient identifier(s) and the occurrences of this *Concept*. The output of the service is then aggregated to a logic value by filtering and/or comparing the output with one or more *Constants*. During this evaluation, a logic value (*true*, *false* or *unknown*) is assigned to every cell for each patient.

The logic values of the *Cells* are combined in *Logical Expressions* using the Kleene logic rules to indicate whether a patient conforms to the set of eligibility criteria defined in a *Group*. Intermediate results trickle up to the top-level '*Included*' *Group*, which collects the overall eligibility result for a specific patient.

A proof-of-concept implementation of an evaluation engine for the proposed clinical trial ontology was developed at the Antwerp University Hospital. This engine connects to several hospital database systems and the EHR system, allowing the evaluation engine to generate lists of eligible candidates from the overall patient population in the hospital.

2.3.4 Candidate ranking

Our implementation of the evaluation engine introduced a ternary logic, leading to a large set of candidates for which eligibility is *unknown* given the current information. Further ranking of these *unknown* candidates provides valuable insights into which candidates could be further investigated to determine their eligibility (e.g. by performing an additional lab test or by asking additional questions). A straightforward approach could assign different costs to each concept *type*, an overall cost per candidate can then be calculated by additively combining the costs of *unknown* concepts. Costs can e.g. be based on the estimated difficulty and price to acquire the necessary information. This approach however does not correctly rank the candidates to determine eligibility. E.g. a rule with 10 concepts in an *or* relation, would evaluate to *true* if any single concept is *true* while the overall cost would be based on all 10 concepts.

We therefore developed a more accurate but computationally more complex approach. For each *unknown* candidate, we simulate the effect of determining the value of an individual *unknown* concept by replacing it once with *true* and once with *false*. This leads to a new overall eligibility of the candidate, which can take any of the following values: T/F, T/U, F/T, F/U, U/T, U/F, U/U (note that T/T and F/F can never occur). We now group all candidates in the following four categories: a) T/F, F/T; b) T/U, U/T; c) F/U, U/F; d) U/U. Eligibility of candidates in category a) is fully known by determining the value of a single concept; for cat. b) inclusion can be determined, but not exclusion; for cat. c) only exclusion can be determined; and for cat. d) no information is gained. Within each category we can now apply a cost ranking based only on the concept that was changed. This procedure can be extended to pairs and sets of concepts where the set of *unknown* values is replaced by all possible *true/false* combinations.

3. DISCUSSION

When specific topic maps for clinical trial eligibility are shared between multiple screening centers, two orthogonal aspects are vital: *specialization* and *generalization*.

Specialization indicates that a topic map can be flexibly adapted to meet the specific ICT infrastructure of an individual screening center. When an individual screening center receives such a clinical trial topic map, it would define a new *Institution* instance that serves as scope on the different associations. The screening center can create the necessary new topics and apply its own scope to associations linked to the existing structure. A screening center could locally define the concept *'diabetes mellitus type II'* based on the internally available data. Some screening centers might base this concept on formal codes such as SNOMED (Systematized Nomenclature of Medicine) or ICD-9-CM, others might directly base it on locally defined lab tests. This specialization allows to adapt the topic map to the local IT infrastructure without loss of information and by keeping the integrity of the original topic map structure intact.

Generalization means that the screening center can indicate which parts of the clinical trial structure are directly queryable on its infrastructure without specialization. An example could be the querying of patient demographic data. Demographic data contains many generic concepts such as name, birthdate, gender, ... that are defined in international
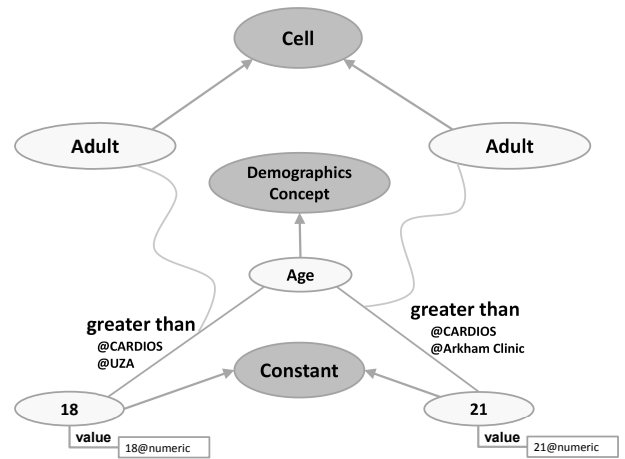

Figure 6: An example adaptation of the concept *Adult* to accommodate for multiple centers with different age thresholds for defining adulthood.

standards, e.g. *open*EHR, HL7 v.3 (Health Level Seven). A screening center that exposes its EHRs via one of these standards can therefore directly answer (part of) a clinical trial where a concept is defined as an *open*EHR query, without the need for specialization of the topic map. This approach enables users to apply these international standards where desired without imposing them.

To illustrate this, we present the following example. The Antwerp University Hospital shares a topic map of the CARDIOS clinical trial with the (fictional) *Arkham Clinic* institution. The clinical trial is restricted to adult participants only. However, the country where Arkham Clinic resides has stricter laws that define adulthood as being at least 21 years old, while adulthood for Antwerp University Hospital (Belgium) is defined as being 18 years old. The topic map representation can then be adapted as in Figure 6 by creating a new Cell 'Adult' that reifies a new *greater than* Association that is scoped with the *Arkham Clinic* topic.

Furthermore, multiple specialized clinical trial topic maps from different centers can be merged into one single topic map. As all structural associations of a topic map are scoped with a topic representing that trial, it is always possible to retrieve the structure of an individual trial among the merged collection of all trials. Merging multiple topic maps provides significant benefits in terms of creation, maintenance and updating when they share the same definitions of *Concepts*, *Cells* or even *Groups*.

The clinical trial ontology encapsulates access to hospital-specific data sources through *Concept* subtypes. While this greatly enhances the reusability of clinical trial topic maps, it still requires additional implementation work when introducing this system in a new institution as it needs to implement the necessary plugins for access to its own data sources. By applying the generalization approaches described in Section 3, institutions would only be required to provide a standards-compliant API to their internal EHR data. Future work includes the incorporation of these standards in the proposed clinical trial ontology.

4. CONCLUSIONS

In this paper, we presented a Topic Maps ontology designed for the evaluation of clinical trial eligibility over multiple screening centers. This ontology enables the creation of formally structured clinical trials topic maps. We showed how these topic maps can be used in an automated evaluation engine to provide a set of eligible clinical trial candidates. Our evaluation engine is capable of both directly linking with existing EHR systems and enables users to use international standards where possible without imposing them. Furthermore, we showed how the use of Topic Maps addresses the goal of sharing eligibility criteria, concepts and formalized clinical trial protocols between multiple institutions and offers both a formalized yet flexible ontology that can be locally extended.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Clinical research filtered query. http://wiki.hl7.org/index.php?title=Product_RR_CRFQ.

[2] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

[3] R. W. Carlson, S. W. Tu, N. M. Lane, T. L. Lai, C. A. Kemper, M. A. Musen, and E. H. Shortliffe. Computer-based screening of patients with HIV/AIDS for clinical-trial eligibility. *The Online journal of current clinical trials*, Doc No 179, Mar. 1995.

[4] D. Damen, R. Pinchuk, and B. Fontaine. Creating Topic Maps Ontologies for Space Experiments. In L. Maicher and L. M. Garshol, editors, *Fifth International Conference on Topic Maps Research and Applications: Linked Topic Maps*. University of Leipzig, 2009.

[5] L. Fallowfield, D. Ratcliffe, and R. Souhami. Clinicians' attitudes to clinical trials of cancer therapy. *European journal of cancer (Oxford, England : 1990)*, 33(13):2221–2229, Nov. 1997.

[6] E. Fink, L. O. Hall, D. B. Goldgof, B. D. Goswami, M. Boonstra, and J. P. Krischer. Experiments on the automated selection of patients for clinical trials. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, pages 4541–4545. IEEE, 2003.

[7] M. Fitting. Kleene's three valued logics and their children. *Fundamenta Informaticae*, 20(1):113–131, Jan. 1994.

[8] L. M. Garshol. Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science*, 30(4):378–391, Aug. 2004.

[9] L. M. Garshol. *Towards a Methodology for Developing Topic Maps Ontologies*, volume 4438 of *Lecture Notes in Computer Science*, chapter 3, pages 20–31. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[10] O. Hassanzadeh, A. Kementsietsidis, L. Lim, R. J. Miller, and M. Wang. LinkedCT: A Linked Data Space for Clinical Trials. Aug. 2009.

[11] I. Horrocks, P. F. Patel-Schneider, and F. Van Harmelen. From SHIQ and RDF to OWL: The Making of a Web Ontology Language. In *Journal of Web Semantics*, volume 1, pages 7–26, 2003.

[12] C. A. Kotwall, L. J. Mahoney, R. E. Myers, and L. DeCoste. Reasons for non-entry in randomized clinical trials for breast cancer: a single institutional study. *Journal of surgical oncology*, 50(2):125–129, June 1992.

[13] O. Lassila and R. R. Swick. Resource Description Framework (RDF) Model and Syntax Specification. 1998.

[14] H. Niederhausen, S. Windisch, and L. Maicher. Generating an Ontology Specific Editor. In M. Popescu and D. L. Stewart, editors, *SEMAPRO 2010, The Fourth International Conference on Advances in Semantic Processing*, Oct. 2010.

[15] J. Niland. ASPIRE: Agreement on Standardized Protocol Inclusion Requirements for Eligibility, 2007.

[16] S. Pepper. The TAO of Topic Maps - Finding the Way in the Age of Infoglut. 2000.

[17] F. A. Thiers, A. J. Sinskey, and E. R. Berndt. Trends in the globalization of clinical trials. *Nat Rev Drug Discov*, 7(1):13–14, Jan. 2008.

[18] S. W. Tu, C. A. Kemper, N. M. Lane, R. W. Carlson, and M. A. Musen. A methodology for determining patients' eligibility for clinical trials. *Methods of information in medicine*, 32(4):317–325, Aug. 1993.

[19] S. W. Tu, M. Peleg, S. Carini, M. Bobak, J. Ross, D. Rubin, and I. Sim. A practical method for transforming free-text eligibility criteria into computable criteria. *Journal of biomedical informatics*, 44(2):239–250, Apr. 2011.

[20] C. Weng, S. W. Tu, I. Sim, and R. Richesson. Formal representation of eligibility criteria: a literature review. *Journal of biomedical informatics*, 43(3):451–467, June 2010.

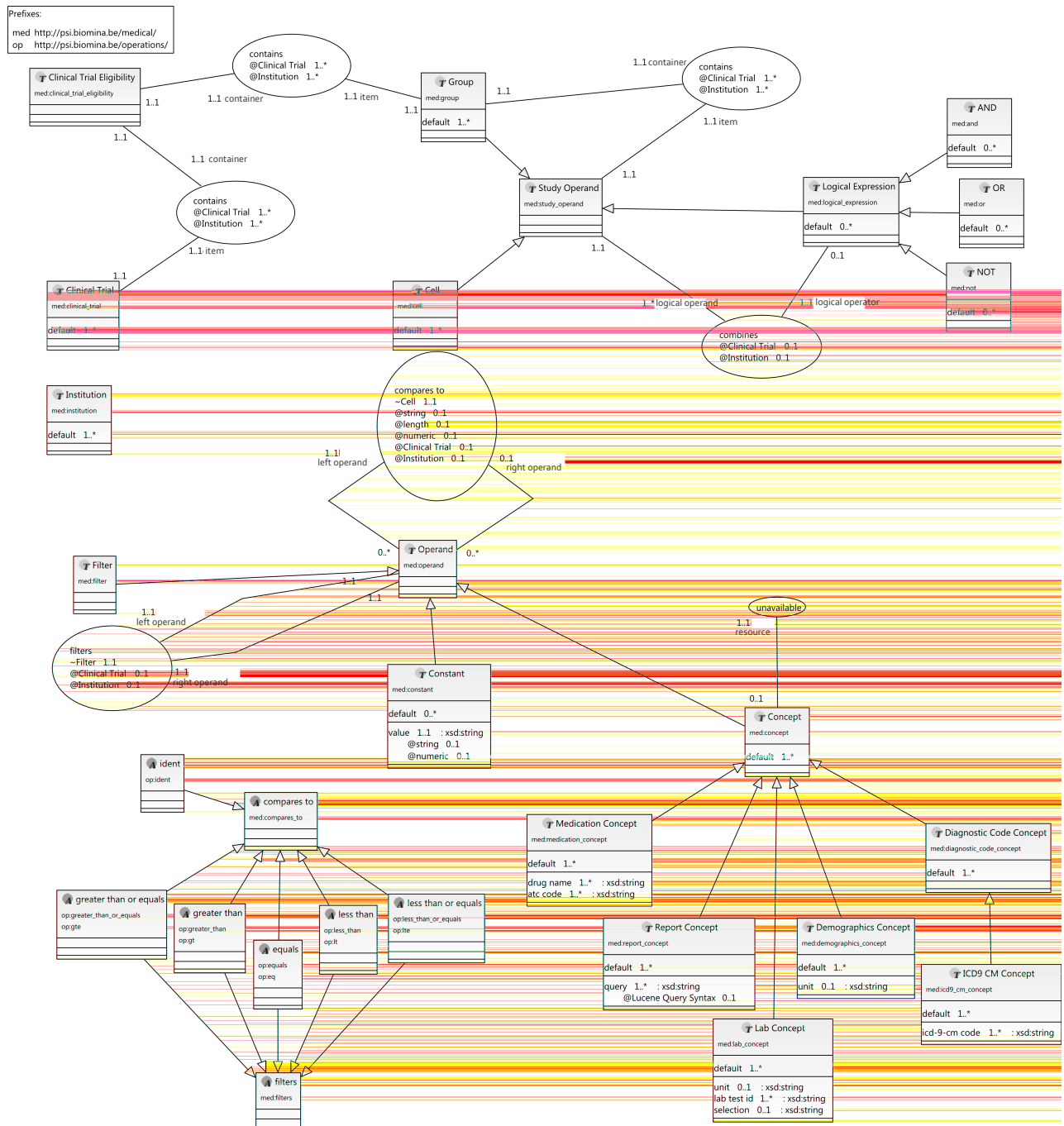## APPENDIX

A. CLINICAL TRIAL ONTOLOGY

Figure 7: Clinical trial Topic Maps ontology.
Graphical representation of the complete ontology for clinical trials. It has been created with Onotoa [14]. A description of the representation format of the ontology figure can be found at the Onotoa website.