SUNY / PRAXIS EMR RESEARCH ON INTEROPERABILITY AND QUERY OF CLINICAL RECORDS

(3/13/08)

JOINT STATE UNIVERSITY OF NEW YORK (SUNY) - PRAXIS EMR RESEARCH ON INTEROPERABILITY AND QUERY OF CLINICAL RECORDS

10/4/2007: This is a documentation excerpt submitted to the NIH by the New York State Center of Excellence in Bioinformatics and Life Sciences (SUNY) in conjunction with Praxis Electronic Medical Records, seeking a research grant to develop this revolutionary interoperability engine. As noted in the paper by the Center below, no other approaches to interoperability based on the standard languages such as SNOMED and MEDCIN alone will be sufficiently valid. This paper also partly explains why those approaches have not yet been successfully implemented. We wholly agree with the Center's approach because it will not affect the freedom of doctors to chart their own way at extraordinary speed, which has been the basic tenet of Praxis all this time. Please also review our other papers on ***Clinical Practice Guidelines and Queries*** *as well as* ***The 3R's*** *by Doctor Clayton Reynolds.*

See SUNY links:

http://www.org.buffalo.edu/RTU

http://www.org.buffalo.edu/RTU/indcollabs.html

…and Praxis Links on practice guidelines and queries:

Praxis Clinical Practice Guidelines

The 3R's

1. ABSTRACT

The Infor-Med Corporation in collaboration with the New York State Center of Excellence in Bioinformatics and Life Sciences proposes to explore how to integrate Referent Tracking and Basic Formal Ontology into the Praxis® EMR. Praxis® is an Electronic Medical Record (EMR) system that allows physicians to manage patient data in accordance with their unique practice management preferences and their insight into what will be the best treatment for their patients rather than by following a structured user interface developed by a software engineer. Drawbacks of this approach, however, are that data gathered by one physician are not automatically comparable with data gathered by another one using a different EMR, and that linking to other information sources is very difficult. This is problematic since the future of medicine lies not only in portability of records as patients traverse the health care system but also in the linking of patient data to other sources of information such as the vast stores of data that are added daily by biomedical research.

The integration of Referent Tracking (RT) and Basic Formal Ontology (BFO) in Praxis® will solve these problems, and will add a level of interoperability to this system making it an even greater asset to physicians in the treatment of their patients. The innovation of the proposal lies in its use of RT and BFO to disambiguate the data elements generated by the Praxis® EMR ***before*** they are linked to the terms from a controlled medical vocabulary such as SNOMED CT® or Medcin. **Disambiguation at this level is an issue that is generally overlooked in prevailing paradigms on EMR keeping. Indeed, attempts to achieve interoperability based on vocabularies alone will, in the Center's estimate, fall short of the scope of data integration needed to deliver the full promise of healthcare information technology. With the technology that Doctor Werner Ceusters and his SUNY team propose to deliver, Praxis® users will be able to enjoy the same degree of freedom in parameterizing and using the system according to their specific needs, yet be fully compatible with national goals of achieving interoperability in healthcare information technology.**

2. SPECIFIC AIMS

Three specific aims will be achieved during Phase I:
Specific Aim 1 Collect in accordance with all HIPAA regulations a set of (similar) case descriptions provided by Praxis® users, and analyze them to assess the extent to which the data within them can be formalized and, by being so, be made interoperable,
Specific Aim 2 Measure to what extent SNOMED CT® terms and concepts are able to cover the terminology used within case descriptions, and
Specific Aim 3 Create a set of requirement specifications for a new component of the Praxis® EMR that will guide physicians during the creation and revision of case descriptions towards creating structured language that is conducive to interoperability.

The requirement specifications will then serve as input for Phase II during which the prototype will be implemented.

3. BACKGROUND AND SIGNIFICANCE

3a. Background

The Referent Tracking Unit of the Center of Excellence in Bioinformatics & Life Sciences

The Referent Tracking Unit (RTU) operates under the auspices of the New York State Center of Excellence in Bioinformatics and Life Sciences (CoEBLS), a powerful collaboration between state and local government leaders, business

leaders and the three major research institutions of Western New York: The State University of New York at Buffalo (UB), Roswell Park Cancer Institute (RPCI) and the Hauptman-Woodward Medical Research Institute (HWI). CoEBLS focuses on merging novel computational approaches and techniques with biomedical sciences to speed the generation of new knowledge.

The mission of the Referent Tracking Unit (RTU) is to carry out fundamental and applied research and software application development with the goal of allowing better use to be made of both (1) data pertaining to particular patients residing in EHRs on the one hand, and (2) patient-independent data of the type that is typically found in biomedical research databases on the other. The work of the RTU is designed to allow biomedical and bioinformatics researchers to exploit the wealth of information that is stored in patient data repositories. At the same time, it is designed to offer clinicians new and higher quality types of evidence for the appropriateness of given diagnoses or therapeutic hypotheses through seamless access to the research data generated by biologists and bio-informaticians.

The director of the RTU, Werner Ceusters, studied medicine, neuro-psychiatry, informatics, and knowledge engineering in Belgium; since 1993 he has been involved in numerous national and European research projects in the area of Electronic Health Records, Natural Language Understanding and Ontology. Prior to his tenure with the RTU, he was Executive Director of the European Centre for Ontological Research at Saarland University, Germany. He is currently Health Sciences Professor connected to the Psychiatry Department of the School of Medicine and Biomedical Sciences, UB, Director of the Ontology Research Group of the New York State Center of Excellence in Bioinformatics and Life Sciences, and coordinator of Bioinformatics for the Health Science Faculties at UB.

Towards semantic interoperability

If there is an issue regarding interoperability, it is not ***whether*** an EMR should generate data that are interoperable but rather ***how*** an EMR should generate data that are interoperable. Here, ‘interoperable data’ means data created and stored in an EMR which are reusable by other software applications in such a way that those other applications can perform the functions for which they were designed with minimum human intervention. The growing national consensus on how an EMR should generate such data is that controlled vocabularies will play a prominent role.

National plans to address the issue of interoperability of healthcare data are beginning to take shape. On April 27, 2004 President Bush signed an executive order[7] creating the position of Health Information Technology Coordinator. The

responsibilities of the Coordinator include directing "… the implementation of a strategic plan to guide the nationwide implementation of interoperable health information technology in both the public and private health care sectors…". In the three years since the signing of this order much work has been completed toward the creation of this strategic plan culminating on May 11, 2007 with a set of three interoperability specifications [8-10] released by the Healthcare Information Technology Standards Panel (HITSP).

The common design in all of the HITSP specifications is that document and messaging standards such as the American Society for Testing and Materials (ASTM) Standard Specification for Continuity of Care Record (CCR) [11] and the Health Level Seven (HL7) Version 3.0 [12] are combined with vocabulary standards such as the International Health Terminology Standards Development Organization's (IHTSDO) Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT®) [13] and the American Medical Association's (AMA) Current Procedural Terminology (CPT®) Fourth Edition (CPT-4) [14].

To succeed in producing interoperable data, these national plans must be implemented in an EMR and for EMRs to be commercially viable, they must satisfy physician requirements of recording all data they deem to be important in a way that is fast, intuitive, and flexible. The use of natural language is well-suited to achieving the latter but comes with the cost of prohibiting the former. Thus, within an EMR, the problem of interoperability comes down to finding a means for allowing physicians to record data in an effective manner and be able to annotate this data with codes from vocabularies without introducing ambiguities at either step.

Vocabularies alone are not enough

Although vocabularies have been in existence for decades, they have thus far not been able to solve problems of semantic interoperability. This is due, not only to shortcomings in the vocabularies themselves, but rather to the incapacity of traditional EMR systems to allow data to be annotated by means of terminologies in ways that do not ***create*** ambiguities. EMRs using standard terminologies generally do no more than record that a patient has ***some*** instance of a sign, symptom, or disease and as a result lose the important ability to refer to the patients' ***unique and particular*** instances of signs, symptoms and diseases. Multiple instances of a symptom type cannot be disambiguated easily from multiple references to a single instance of a symptom type. For example, a patient who experiences multiple syncopal episodes will have a list of chief complaints that is no different, at least not to a software agent designed to process elements in the list, from the problem list of a patient who revisits his or her general practitioner for one and the same syncopal episode multiple times.

Another result of this annotation method is the loss of data that occurs as a condition is charted through its evolution over time. A benign intestinal polyp

discovered during a colonoscopy which later becomes malignant will be documented with two entries in the patient's record, each having a distinct code from a vocabulary but neither preserving the information that the same entity is being referred to by both.

3b. Significance

The analysis of case descriptions to be performed in Phase I will identify those constructions within case descriptions that either enhance or impede interoperability. **As Doctor Ceusters and his team at the Center state: "ultimate goal is to achieve a level of interoperability yet to be thought of by Praxis competitors. Avoiding the weaknesses that exist in current mainstream plans for interoperability, our work will make the Praxis EMR the leader in EMR interoperability."**

4. PRELIMINARY STUDIES

- **4b. The Cassandra Syntax**

In [16], Ceusters et al described the Cassandra syntactic-semantic tagging system. This system transforms sentences expressed in natural language into a structured representation that is independent of the subtleties of linguistic structure that underlies the way natural languages work. The structured representation eliminates sources of ambiguity thereby improving subsequent computational processing for information retrieval, automated translation and language understanding. In [17], principles behind controlled language design and use are explained through a detailed study of the inconsistencies and ambiguities that arise when interpreting SNOMED procedure terms in the framework of medical information and it is shown that most of them can be explained as a violation of sound term-formation principles. In [18], the use of Cassandra in mediating between the two controlled of SNOMED and GALEN is described.

The Ceusters experience with the Cassandra syntax will allow the Center to relate the structure of a sentence to its meaning in ways that have proved to be successful. The syntax provides a tagging scheme that can be used in conjunction with the existing bracketing method to recapture lost information and to ensure correct coding of terms.

- **4c. Basic Formal Ontology**

Basic Formal Ontology is an ontological theory of the general types of entities that exist. Within BFO, the main subdivision of entities is between universals and particulars. Universals are the natural kinds in reality denoted by the general

terms of science. Particulars, on the other hand, are the instances of such universals and are the individuals we encounter in our experience of the world. Universals are what similar individuals have in common; they are the invariants in reality.

An important distinction among particulars made by BFO is based upon whether or not they have temporal parts, that is, on whether or not at any moment of time an entity is fully present or is instead only partially present. The former type of entity is a *continuant* and the latter an *occurrent*. An example of a continuant is the fracture of a patient's left femur and an example of an occurrent is the process through which that fracture is healed.

A subdivision of continuants is that between *independent* and *dependent* entities. Molecules and cells are examples of independent entities. The shapes of those molecules and cells are dependent entities as these require the former in order to exist (in an ontological sense of 'require' that is different from a biological sense of 'require' used when we say that organisms require food or oxygen). A patient's left femur is an independent continuant – there is no other particular on which it depends in this ontological sense. The fracture of a patient's left femur, in contrast, depends ontologically upon the femur and is therefore a dependent continuant. Occurrents are dependent entities exclusively, relying always upon at least one continuant as their bearer.

In this project, the principles of BFO will be used to categorize terms by the different upper-level ontological kinds that subsume the particulars which they denote. This categorization will in the first place assist in identifying the structure of sentences and secondly it will allow us to develop guidelines to prevent physicians using our component to construct sentences that contain ontological errors, or ambiguities that are common when using natural language. For example, processes are often confused with the particulars that bear them. Avoiding these confusions with the assistance of the principles of BFO will be crucial to achieving interoperability.

4d. The Relation Ontology

The Relation Ontology is a theory of the primary ontological relations that hold between entities. The relations are ontological in the sense that they exist between entities independently of our ways of gaining knowledge about such entities and independently of our ways of representing or processing such knowledge. That an infection *causes* a patient's fever is an ontological relation while that a fever *is evidence of* an infection is not.

The relations included in the Relation Ontology are: foundational relations (is_a, part_of), spatial relations (located_in, contained_in, adjacent_to), temporal relations (transformation_of, derives_from, preceded_by), and participation

relations (has_participant, has_agent). The Relation Ontology provides rigorous definitions of these relations as they apply to the entities referred to in BFO.

In this project, RO will be used to categorize relations between terms by the upper-level ontological relations which subsume them. This, as with the application of the principles of BFO, will serve a dual purpose: the first being the identification of the structure of sentences and the second being to aid in guiding physicians away from constructing sentences that contain errors or ambiguities that are common in natural language. An example is that of the use of oblique reference as occurs in the second of the two following sentences: “Patient has [2 inch] laceration of [left | right] foot.” “The loss of tissue is [none | minimal | moderate | substantial].” Understanding that tissue loss is related to a cause will result in the second sentence being flagged as a structure which needs to be supplemented with a link to the term in the first sentence in order to preserve its intended meaning.

4e. Referent Tracking

The Referent Tracking (RT) paradigm was introduced in 2005 [1] and its requirements and infrastructure were detailed in 2006 [19]. Its goal is to reduce the ambiguous references within EMRs by introducing globally unique and singular identifiers, called IUIs, for the particular entities currently referred to by means of general terms taken from a terminology. Thus, not only patients and physicians are uniquely identified, but so also are the patients’ diseases, the signs and symptoms they exhibit, and the treatments administered. Management of IUIs is performed by a referent tracking system (RTS) [20] designed to deliver services to EMRs installed at separate locations in a health care network. The RTS architecture provides the capability for unambiguous reference to any entity referred to within the system even as information pertaining to this entity is recorded by distinct health care providers in distinct health care settings and potentially using distinct EMR applications.

In turn, Referent Tracking is framed against the background of BFO and the RO [21] of the Open Biomedical Ontologies (OBO) Consortium. The OBO is an informal, voluntary initiative of developers of cross-linked biomedical ontologies and is representative of a general trend on the part of the National Institute of Health, Food and Drug Administration, and other bodies to consolidate ontology-based standards for the communication and processing of biomedical data. The Consortium produces ontologies based on a common set of principles designed to assure that they are intelligible to their users, stable, interoperable, and support logic-based reasoning. The usefulness of this methodology is evidenced by the clinical uses that are starting to be made of the Gene Ontology[22], the most widely-used of the OBO ontologies. Aligning to the top-level ontologies of the Basic Formal Ontology and Relation Ontology is central to the principles of the Consortium.

In this project, the principles of Referent Tracking will be used to identify the structures that physicians use to refer to the particulars of a patient's condition and treatment within a case description of a disease type. The sentence "Patient has [greenstick | spiral | transverse | comminuted | open] fracture of [left | right] [femur]" that might occur in the examination section of the encounter note refers to the same particular as the sentence "Manipulation of fracture occurred after treatment with [hematoma block | intravenous sedation]." that might occur in the treatment section. While the continuity of reference between the two sentences is no doubt clear to a human reader of the entire note, it would be lost to a software agent. Providing physicians with a guide to link references to the same particular would not only preserve this important information but would also produce a time-saving way of annotating all occurrences of the term that referred to the particular with the annotation of one of them.

5. RESEARCH DESIGN AND METHODS

5a. Specific Aim 1

Collect in accordance with all HIPAA regulations a set of (similar) case descriptions provided by Praxis® users, and analyze them to assess the extent to which the data within them can be formalized and, by being so, be made interoperable.

5a1. Rationale and Objectives

Fortunately, because of the manner in which the Concept Processor encourages physicians to document case descriptions consistently and the bracketing mechanism already in use in the Praxis® EMR, there is a simpler, cost-effective and more reliable way of achieving the goal of Semantic interoperability goal. The objective of this specific aim will be to gain a full understanding of the sentence structures that physicians employ to describe their patients' conditions. The variations will be documented with a description and an assessment of the degree to which the given structure is amendable to formalization.

Analysis of the case descriptions will begin by collecting them from current users. The Praxis® EMR contains a straightforward mechanism through which these can be exported. The data will be transferred in accordance with all HIPAA requirements. The content of these descriptions is illustrated in both the History of Present Illness screenshot in Figure 1 and in the Review of Systems screenshot in Figure 2 above. Once the descriptions have been collected we will categorize each by the condition it covers. The categorization will be performed to facilitate discovery of how the same symptoms and signs are described by different physicians: we expect that different physicians will tend to examine and treat the same conditions in similar, if not in identical, ways, but also that the way they document this, will vary.

5a2. Design

5a3. Methods and Materials

There are two measures of importance for the analysis of case description content: the first is what constitutes a variation between structures and the second is the level of formalization that a structure satisfies.
Variation may be found in:

1) the grammar of expressions.

2) the ontological reference of expressions, as in the case of the two sentences “Patient is [hypotensive | hypertensive]” and “Patient [has | does not have] hypertension” where the reference of the first is a disposition of the patient’s cardiovascular system (a dependent continuant) and the second refers to a disease process (an occurrent).

3) the relation between terms.

4) the use of implicit references. We do expect to discover during the analysis of case descriptions numerous subtypes of variations within each top level categories. Each subtype will be described as it is discovered.

The level of formalization of a structure type will be defined by the amount of semantic information it contains (terms, relations between terms, explicit reference) that can be demarcated using an extension of the bracketing mechanism currently available in Praxis®, and this along the lines of the Cassandra syntax.

The outcome of this task will be a list of grammatical structure types employed by Praxis® EMR users. The list will contain 1) a schema of the structure, 2) a schema of the formalization of the structure, and 3) an example of a sentence exemplifying the structure.

5a4. Results

5b. Specific Aim 2

Measure to what extent SNOMED CT® terms and concepts are able to cover the terminology used within case descriptions

5b1. Rationale and Objectives

While the case descriptions are analyzed in accomplishment of Specific Aim 1 we will also be able to match terms identified within those descriptions to terms in the SNOMED CT® vocabulary. This provides us with a test of two important

factors: 1) what percentage of terms identified by our analysis match to SNOMED CT® terms and 2) if two different structures denote the same particular, do they both match to the same concept in SNOMED CT®.

5b2. Design

Terms extracted from case descriptions will be entered into a database table. An existing term matching algorithm (e.g. MetaMap Transfer[32]) will compare the case description terms with SNOMED CT® and if a match is found, the applicable SNOMED CT code will be entered into a column in the case description term database. While performing this task, we will keep track of:

- terms that were automatically mapped, and further, after manual inspection, whether the mapping was correct

- terms that could not be mapped automatically, but for which manual inspection revealed a mapping SNOMED_CT concept
- terms for which no SNOMED CT concept exists.

This process will be an iterative one, where after each run of the matching process failures in matching will be studied to determine the source of the failure. This will allow us to refine our structure schemas to maximize success rates and to identify certain structures that need to be avoided in the construction of case descriptions.

5b3. Methods and Materials

The list of terms to be used in the accomplishment of Specific Aim 2 is one of the by-products of Specific Aim 1. Starting from a set of case description sentences, the structure and location of terms within a sentence will be identified in order to determine the extent to which a given expression can be formalized.
An existing term matching algorithm (e.g. MetaMap) will be used to automate the matching process between the case description terms and the terms from SNOMED CT®.

5b4. Results

The outcome of this task will be a set of statistics on the success and failure rates of matching terms from case descriptions to terms from SNOMED CT®. Four metrics will be compiled 1) percentage of correctly matched terms, 2) percentage of correctly unmatched terms (i.e. no matching term in SNOMED CT®), 3) percentage of incorrectly matched terms (term returned from SNOMED CT® is not synonymous with case description term), and 4) percentage of incorrectly unmatched terms (i.e. a matching term from SNOMED CT® exists but was not returned). These metrics will allow us also to quantify recall and precision [33].

5c. Specific Aim 3

Create a set of requirements specifications for a new component of the Praxis® EMR that will guide physicians during the creation and revision of case descriptions towards creating structured language that is conducive to interoperability.

5c1. Rationale and Objectives

Through specific Aims 1 and 2, we will understand which types of expressions currently employed by Praxis® users are conducive to generating interoperable data.

5c2. Design

The requirements specifications to be created will contain 7 sections: 1) Product Vision – an overview of the purpose of the new component, 2) Use cases – descriptions of the real world situations that will be encountered and the components responses to them, 3) User Interfaces – a description of all the user interfaces of the new component including a navigation path between them, 4) Software interfaces – a description of all the existing components of the Praxis® EMR that will need to be modified during the creation of the new component, 5) Performance requirements – description of how the Praxis® EMR is expected to function during execution of the new component, 6) Database changes – a description of the entire set of changes that will new to be made in order to implement the new component, 7) Testing requirements – a description of the ways in which the Praxis® EMR will need to be tested in order to assure that the new component matches its requirements and the regression testing that needs to be performed to assure that the new component does not interfere with any existing Praxis® EMR functionality.

1. Ceusters, W., Smith, B., *Tracking Referents in Electronic Health Records*, in *Connecting Medical Informatics and Bio-Informatics: Proceedings of MIE 2005 The XIXth International Congress of the European Federation for Medical Informatics*, G.A. Engelbrecht R., Lovis, C., Editor. 2005, IOS Press: Amsterdam. p. 71-76.
2. Grenon, P., Smith, B., Goldberg, L., *Biodynamic Ontology: Applying BFO in the Biomedical Domain*, in *Ontologies in Medicine*, D.M. Pisanelli, Editor. 2004, IOS Press: Amsterdam. p. 20-38.
3. Gruber, T.R., *Toward principles for the design of ontologies used for knowledge sharing.* International Journal of Human-Computer Studies, 1995. **43**(5-6): p. 907-928.
4. *HIMSS Davies Awards*. 2007 [cited 2007 July 23]; Available from: http://www.himss.org/asp/davies_primarycare.asp.
5. Smith, B., J. Williams, and S. Schulze-Kremer, *The Ontology of the Gene Ontology*, in *AMIA Annual Symposium Proceedings*. 2003. p. 609-613.
6. Weed, L., *Medical Records That Guide and Teach.* New England Journal of Medicine, 1968. **278**: p. 593-600.
7. Smith, B., *Beyond concepts: ontology as reality representation*, in *Proceedings of the third international conference on formal ontology in information systems (FOIS 2004)*. 2004, IOS Press: Amsterdam. p. 73-84.
8. *HITSP Electronic Health Records Laboratory Results Reporting Interoperability Specification*. 2007 [cited 2007 July 23]; Available from: http://www.ansi.org/standards_activities/standards_boards_panels/hisb/hitsp.aspx?menuid=3.
9. *HITSP Biosurveillance Interoperability Specification*. 2007 [cited 2007 July 23]; Available from: http://www.ansi.org/standards_activities/standards_boards_panels/hisb/hitsp.aspx?menuid=3.
10. *HITSP Consumer Empowerment Interoperability Specification*. 2007 [cited 2007 July 23]; Available from: http://www.ansi.org/standards_activities/standards_boards_panels/hisb/hitsp.aspx?menuid=3.
11. *American Society for Testing and Materials International*. 2007 [cited 2007 July 23]; Available from: http://www.astm.org.
12. *Health Level Seven Version 3.0*. 2007 [cited 2007 July 23]; Available from: http://www.hl7.org/.
13. *International Health Terminology Standards Development Organization*. 2007 [cited 2007 July 23]; Available from: http://www.ihtsdo.org.
14. *American Medical Association Current Procedural Terminology*. 2007 [cited 2007 July 23]; Available from: http://www.ama-assn.org/ama/pub/category/3113.html.
15. CEN/ISSS eHealth Standardization Focus Group, *Current and future standardization issues in the eHealth domain: Achieving interoperability*. 2005, CEN/ISSS.
16. Ceusters, W., Waagmeester, A., DeMoor, G. *Syntactic-semantic tagging convention for a medical treebank: The CASSANDRA approach.* in *Proceedings MIC '97*. 1997. The Netherlands: Veldhoven.
17. Ceusters, W., Steurs, F., Zanstra, P., Van Der Haring, E., Rogers, J., *From a Time Standard for Medical Informatics to a Controlled Language*

for Health. International Journal of Medical Informatics, 1998. **48**(1-3): p. 85-101.
18. Ceusters, W., Rogers, J., Consorti, F., Rossi-Mori, A., *Syntactic-semantic tagging as a mediator between representations and formal models: an exercise in linking SNOMED to Galen.* Artificial Intelligence in Medicine, 1999. **15**: p. 5-23.
19. Ceusters, W., Smith, B., *Strategies for referent tracking in electronic health records.* Journal Biomedical Informatics, 2006. **39**(3): p. 362-378.
20. Manzoor, S., Ceusters, W. *Referent Tracking System*. [cited March 9, 2007]; Available from: http://sourceforge.net/projects/rtsystem.
21. Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C.J., Neuhaus, F., Rector, A., Rosse, C., *Relations in Biomedical Ontologies.* Genome Biology, 2005. **6**(5): p. R46.
22. Sjoblom, T., Jones, S., Wood, L., Parsons, D.W., Lin, J., Barber, T., Mandelker, D., Leary, R., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S., Willis, J., Dawson, D., Willson, J., Gazdar, A., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B., Bachman, K., Papdopoulos, N., Vogelstein, B., Kinsler, K., Velculescu, V., *The Consensus Coding Sequences of Human Breast and Colrectal Cancer.* Science, 2006. **314**(5797): p. 268-274.
23. Rudnicki, R., Ceusters, W., Manzoor, S., Smith, B.,, *What Particulars are Referred to in EHR Data? A Case Study in Integrating Referent Tracking into an Electronic Health Record Application, in AMIA 2007.* 2007. **(Under review)**.
24. Manzoor, S., Ceusters, W., Rudnicki, R., *A Middleware Approach to Integrate Referent Tracking in EHR Systems, in AMIA 2007.* 2007. **(Under review)**.
25. Elkin, P.L., Brown, S.H., Husser, C.S., Bauer, B.A., Wohner-Roedler, D., Rosenbloom, S.T., Speroff, T., *Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists.* Mayo Clinic Proceedings, 2006. **81**(6): p. 741-748.
26. Shah, N.H., Rubin, D.L., Supekar, K.S., Musen, M.A., *Ontology-based annotation and query of tissue microarray data.* AMIA Annual Symposium Proceedings, 2006: p. 709-713.
27. Chiang, M., Casper, J., Cimino, J., Starren, J., *Representation of ophthalmology concepts by electronic systems Adequacy of controlled medical terminologies. .* Ophthalmology, 2005. **112**(2): p. 175-183.
28. Dykes, P.C., Currie, L.M., Cimino, J.J., *Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway.* Journal Biomedical Informatics, 2003. **36**: p. 313-325.
29. Ceusters, W., Smith, B., *Ontology and Medical Terminology: Why Description Logics are not enough.* Proceedings of the Conference Towards an Electronic Patient Record (TEPR 2003), 2003.
30. Ceusters, W., Smith, B., Kumar A., Dhaen C., *Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be*

Detected?, in *Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies, Rome October 2003*, D.M. Pisanelli, Editor. 2004, IOS Press.

31. Ceusters, W., Smith, B., Kumar A., Dhaen C. *Ontology-Based Error Detection in SNOMED CT®*. in *Proceedings of Medinfo*. 2004.
32. *MetaMap Transfer (MMTx)*. 2007 [cited 2007 July 23]; Available from: http://mmtx.nlm.nih.gov.
33. Ceusters, W. *Towards a Realism-Based Metric for Quality Assurance in Ontology Matching*. in *Formal Ontology in Information Systems*. 2006. Baltimore, MD: IOS Press Amsterdam.

SNOMED CTSNOMED CT