

# What Particulars are Referred to in Electronic Health Record Data? A Case Study in Integrating Referent Tracking into an EHR Application

Ron RUDNICKI<sup>1</sup>, Werner CEUSTERS<sup>1,2</sup>, Shahid MANZOOR<sup>1</sup>, Barry SMITH<sup>1,2</sup>

<sup>1</sup> Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, NY, USA

<sup>2</sup> National Center for Biomedical Ontology, University at Buffalo, NY, USA

## Abstract

*Referent Tracking (RT) advocates the use of instance unique identifiers to refer to the entities comprising the subject matter of patient health records. RT promises many benefits to those who use health record data to improve patient care. To further the adoption of the paradigm we provide an illustration of how data from an EHR application needs to be decomposed in order to make it accord with the tenets of RT. We describe the ontological principles on which this decomposition is based in order to allow integration efforts to be applied in similar ways to other EHR applications. We find that an ordinary statement from an EHR contains a surprising amount of “hidden” data that are only revealed by its decomposition according to these principles.*

## Introduction

The Referent Tracking (RT) paradigm was introduced in 2005<sup>1</sup> and its requirements and infrastructure were detailed in 2006.<sup>2</sup> The goal of the paradigm is to reduce ambiguous reference within the electronic health record (EHR) by introducing globally unique singular identifiers, called *IUIs*, for the particular entities currently referred to by means of general terms taken from a terminology such as SNOMED CT. Not only should patients and physicians be uniquely identified, on this paradigm, but so also should the patients' diseases, the signs and symptoms they exhibit, and the treatments administered. Management of *IUIs* is performed by a referent tracking system (RTS) designed to deliver services to EHR applications installed at separate locations in a health care network.<sup>3</sup> 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 EHR applications.

RT offers a novel approach to data annotation. The theoretical usefulness of globally unique and singular identifiers for individual entities – rather than for “concepts” – was introduced by W. Kent in 1978,<sup>4</sup>

but this idea has never been implemented. It is thus no surprise that, while the ontological framework on which RT is based is gaining acceptance in wider circles of the biomedical community, there is at this time no research on RT itself outside our group.

Ambiguous references in current EHRs create obstacles for example to efforts designed to establish regional health information networks because of the need to determine whether multiple references to a given condition in different portions of a distributed record provide evidence of multiple separate instances of that condition or of multiple observations of the same instance. RT solves this problem even if entities change in type from one time to the next. The statements: “X has a dysplastic nevus at time  $t_1$ ” and “X has a melanoma at time  $t_2$ ” have insufficient content, as they stand, for us to be able to discern if the same entity is referred to in both. When data is annotated as prescribed by RT, however, then this allows entities to be followed as they evolve over time. Tracking change through time holds much promise for applications in domains such as post-marketing surveillance and to determine patient outcomes. The same facility allows us to keep track of an entity as our knowledge of it evolves over time.

## Objectives

One challenge ahead lies in furthering the adoption of the paradigm by developers of EHR applications. To meet this challenge, we have begun the process of integrating RT into commercial EHR applications. A part of this process is an analysis of the extent to which the data collected by an EHR application needs to be reformulated to be compatible with the requirements of RT, namely that the particulars assigned a *IUI* (for ‘instance unique identifier’) are instances of the kinds included in Basic Formal Ontology (BFO).<sup>5</sup> The report we provide here illustrates this analysis, and is intended as an educational tool providing guidance on how to conduct the first stage of a full integration plan.

## Materials

The EHR application on which we conducted our analysis is MedtuityEMR produced by Medtuity Inc.

MedtuityEMR is used by providers of both primary and secondary care. It is a Client-Server application developed in C++, which can be run on either Windows 2000 or Windows XP Professional. The database used by the application is the Microsoft SQL Server Desktop Engine.

MedtuityEMR enables a user to generate a fully readable, highly detailed progress note using only point-and-click controls as input. This is done by creating a multitude of control types of which many are built up from one of 4 basic types (Checkbox, Radio Buttons, Checklist, and Number Box) and whose instances are used by clinicians to document the patient encounter. The output of the controls is formed by merging the input from the user into a predefined parameterized sentence. This design is well suited to the integration of RT as each control in the application has a finite set of possible statements as output. Each of these statements can be linked to an RT compatible reformulation during design-time rather than at run-time. MedtuityEMR stores the data that result from manipulating the controls in a compressed XML file having a structure roughly equivalent to that of a SOAP (Subjective, Objective, Assessment, Planning) note.

To demonstrate how data from MedtuityEMR can be made RT compatible, we selected a subset that contained a reasonable level of complexity, is still of manageable proportions, and representative for a significant portion of the application's full data set.

Since there is no qualitative difference between the data captured by the simple and more complex versions of a control type, choosing a simple control (Figure 1) upon which to illustrate our analysis is sufficient, we believe, to accomplish our educational goals. Our choice was the control from the Fracture-femur disease model that is used to enter information on the strength of flexions of the feet.

## Methods

As RT requires its globally unique identifiers to refer only to spatiotemporal particulars (instances), its integration into an EHR application will sometimes require expanding single data elements from an EHR into several data elements. This expansion is necessary in order to make explicit all of the references that an EHR data element contains only implicitly under current paradigms which focus on what are called *concepts*. The expansions that are required follow the *ontological* – in contrast to *biological* – dependency relations that hold between the various types of particulars as described in BFO,

		0	1	2	3	4	5	
<input type="checkbox"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	foot plantar flexio
<input type="checkbox"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	foot dorsi flexion
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	great toe extensi
<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	foot inversion
<input checked="" type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	foot eversion

**Figure 1. MedtuityEMR ‘6-check’ checklist control with measures of strength entered for flexions of a patient’s feet.**

and that, as explained further down, lead to the distinction of the three types of particulars relevant for our purposes, namely: (1) independent continuants (e.g. John Smith), (2) dependent continuants (e.g. John Smith’s left femur fracture), and (3) occurrents (e.g. the healing of John Smith’s left femur fracture).

Data elements which refer directly to independent continuants such as persons and their body parts require no expansion. Elements which refer to other types of particulars such as weights, blood pressures and measurement acts themselves, do require expansion so that all of the particulars on which the particulars they refer to depend are explicitly mentioned. This requirement is meant to ensure that there are no dangling references within the RTS: if the RTS stores a reference to a fracture, it must also store a reference to the bone that is fractured.

## Basic Formal Ontology (BFO)

Within BFO, the main subdivision among particulars 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*.

A subdivision of continuants (but not occurrents) is that between *independent* or *dependent* entities. An independent entity is for example a molecule or a cell. A dependent entity is for example the shape of a molecule or cell. The latter require the former in order to exist (in an ontological sense of ‘require’ that is different from what is involved for example when we say that organisms require food or oxygen). John Smith’s left femur is an independent continuant – there is no other particular on which it depends in this ontological sense. The fracture of John Smith’s left femur, in contrast, depends ontologically upon another continuant particular: this left femur itself. Each of these distinctions among entities is mutually exclusive and pair-wise disjoint. They yield a total of 4 distinct categories of particulars. But since all

occurrent particulars are dependent entities (they all require one or more independent continuants which serve as their bearers) we are left with just 3 categories: dependent and independent particulars on the one hand, and occurrents on the other.

### Referent Tracking

The first step in making an EHR application RT compatible is to make an analysis of how current data from the EHR application need to be restructured. To accomplish this we must complete, for each *type* of assertion in the EHR, the following tasks (based upon the distinctions amongst entities as described in BFO and on the needs which EHR must serve e.g. in providing traceable liability):

1. identify the particulars to which reference is made in the assertion,
2. identify the relations which are stated to hold between the particulars,
3. identify the universals of which the particulars are instances,
4. identify any concepts or terms with which the particulars are annotated,
5. determine whether the assertion consists of a negative finding,<sup>6</sup> and
6. identify the association of a customary name to a particular.

RT requires further information about the state of affairs referred to by an assertion to be expressed by means of one of the following types of statements:

1. the assignment of an IUI to a particular (e.g. that #321 stands proxy for John Smith and #7865 for John Smith's left femur),
2. the description that at the indicated time a certain relationship holds between particulars (e.g. that #7865 is a part of #321, requiring also that "is a part of" is described in a BFO compatible relationship ontology),
3. the description that at an indicated time a particular is an instance of a given universal (e.g. that #7865 isa femur),
4. the annotation of a particular with a code from a concept-based system (e.g. that #7865 may be annotated with the SNOMED CT codes "182060005" or "T-12739"),
5. the description of a negative finding (e.g. that #321 lacks a left femur, i.e. that the time in question is after #7865 broke and before the pieces had grown back together),<sup>6</sup>
6. the association to a particular of a customary name (e.g. that #321's name is 'John Smith'), and

7. the meta-description of a statement, namely, that the statement has been added to the RTS by a specific agent at a given time.<sup>7</sup>

### Results

The data-entry control that we are using as our example (Figure 1) can generate, depending on how it is manipulated by the clinician using it for data entry, up to 10 sentences. In the state shown in Figure 1, the control would generate the following sentences which then are stored in that form by MedtuityEMR in the patient's EHR: "*The patient's strength of right foot plantar flexion is 3/5; strength of left foot plantar flexion is 4/5; strength of right foot dorsi flexion is 3/5; strength of left foot dorsi flexion is 4/5; strength of bilateral great toe extension is 4/5; strength of right foot inversion is 1/5; strength of left foot inversion is 4/5; strength of right foot eversion is 1/5; strength of left foot inversion is 4/5.*"

Each sentence contains, obviously, references to multiple particulars. MedtuityEMR, however, only assigns to the entire data element generated by the control one globally unique identifier which is formed through the concatenation of (i) the identifier it assigns to the patient session during which the control is used with (ii) the identifier it assigned to the control itself. Note that (ii) is the same independently of the patient and session involved. Such a concatenated identifier does not qualify as a IUI for an entity on the side of the patient. Rather, it is as if the identifiers for the various individual particulars are "hidden" in the sentences generated by the control in a way which will cause problems when these sentences are used for reasoning, and may even prevent reasoning from occurring at all.

For the purposes of this paper, we limit our analysis to the first statement which is '*The patients strength of right foot plantar flexion is 3/5*'. We interpret this as being elliptical for: '*The measurement of the strength of the patient's right foot plantar flexion yielded a value of 3 on a scale from 0 to 5.*'

The particulars and associated BFO categories explicitly referred to by this sentence are:

- P1: the patient's *act of right foot plantar flexion* – Occurrent
- P2: *the act of giving counterforce* to P1 – Occurrent
- P3: *the assessment* that the equality of forces with which P1 and P2 are applied justifies a score of 3/5 – Occurrent

Tracing the dependency relations of these particulars reveals the particulars that are implicitly referred to:

- P4: *the examiner* who performed P3 – Independent Continuant
- P5: *the patient's right foot plantar muscle group* – Independent Continuant
- P6: *the disposition* of the patient's right plantar muscle group to plantar flex the patient's right foot with a certain strength – Dependent Continuant
- P7: *the patient* – Independent Continuant

The relationships (taken from the OBO Relation Ontology<sup>8</sup>) that obtain between these particulars are:

- R1: P7 (the patient) **has part** P5 (his right foot plantar muscle group)
- R2: P6 (the disposition of the patient's right plantar muscle group) **inheres in** P5 (his right foot plantar muscle group)
- R3: P5 (the patient's right foot plantar muscle group) **participates in** P1 (the patient's act of right foot plantar flexion)
- R4: P7 (the patient) **is agent in** P1 (the act of right foot plantar flexion)
- R5: P6 (the disposition of the patient's right plantar muscle group) **is realized in** P1 (the act of right foot plantar flexion).
- R6: P3 (the assessment of equality) **is preceded by** P1 (the patient's act of flexion) and P2 (the examiner's act of giving counterforce);
- R7: P4 (the examiner) **is agent in** P2 (the act of giving counterforce to p1)
- R8: P4 (the examiner) **is agent in** P3 (the assessment of equality of the forces with which P1 and p2 are exercised).
- R9: the force with which P1 (the patient's act of plantar flexion) is exercised **is equal to** the force with which P2 (the examiner's act of giving counterforce) is exercised (and is expressed by the score of 3/5)

Finally, for each particular, it must also be specified what universals they instantiate. This must be done at that level which qualifies the universals as instantiating particulars of one of the three categories that indicate whether or not an entity is dependent. This led to four universals, all taken from BFO: Process (occurrent), Object (independent continuant), Disposition (dependent continuant), and Object Aggregate (independent continuant).

The instantiations of these universals are then:

- I1: P1 **is-instance-of** Process
- I2: P2 **is-instance-of** Process

- I3: P3 **is-instance-of** Process
- I4: P4 **is-instance-of** Object
- I5: P5 **is-instance-of** ObjectAggregate
- I6: P6 **is-instance-of** Disposition
- I7: P7 **is-instance-of** Object

So in this case, making the single statement “The patient's strength of right foot plantar flexion is 3/5” from the MedtuityEMR application compatible with the requirements of RT will require translating it into a set of 23 more detailed statements.

## Discussion

The process of expanding a data element such as is illustrated in Figure 1 to make explicit all of the implicit references to particulars that it may contain can be described in a few steps:

- 1) Identify all the particulars that are explicitly referred to by the element in question.
- 2) For each entity determine its BFO category.
- 3) Independent continuants require no further expansion. If an entity is a dependent continuant, identify the independent continuant on which it depends. If an entity is an occurrent, identify the continuants which participate in it.
- 4) Repeat steps 2) and 3) as required.

These steps need to be performed only once: when the EHR system is integrated with a RTS. Though simple to state, their application can be anything but simple. The ontological distinctions and analyses on which RT is based need to be kept in mind all the time if errors are to be avoided. Dispositional qualities like *strength*, for instance, inhere only in continuants and not in occurrents. This guides the assignment of the patient's strength to his muscles rather than to his flexion act. Strength is a disposition to act in a certain way. If strength were assigned to the acts in which that disposition is realized, then a medical record database would contain references to multiple strengths, one for each *particular* act. This would hinder attempts to retrieve information on how a patient's strength changed over time.

The reader will perhaps have wondered why the patient's right foot was not included in our analysis. There can be no question that the right foot participates in every act of right foot plantar flexion and thus should have been identified at step 3 in our list above. To this we answer that analysis must stop somewhere and here judgment must be exercised (in the same way that it is exercised when deciding what to record in an EHR under current paradigms). Using step 3 unrestrictedly would have led us to include every anatomical feature of the lower leg. We

deemed the patient's right foot to be a passive participant in the mentioned act and thus to be of diminished significance for the description of the finding. The same sort of question can be asked of our decision to include the right plantar muscle group but not to include the 3 individual muscles that comprise it. Here again the finding in question concerned the strength of the muscles acting as a group and consequently the individual muscles of which that group is comprised have diminished significance and need not be listed in the expansion of the finding. Clearly, however, these separate muscles may need to be included in more detailed analyses, for example where their specific modes of operation are affected differentially through some lesion.

By choosing to interpret the data element from MedtuityEMR as an assertion describing an assessment on the part of an examiner of an act of measurement of a quality of a muscle group of a patient, we took the risk of making the integration of an EHR application with RT appear unwieldy. Once that choice was made, unpacking what had appeared to be a simple data element into its component parts revealed a surprising level of complexity. An alternative interpretation of the data element would have been as an entity-attribute-value statement of the form 'right plantar muscle group strength - 3/5'. Following the example of the Phenotype and Trait Ontology (PATO) group,<sup>9</sup> this statement can be simplified into an entity-quality statement. Under this interpretation, there would be two particulars –the muscle group, and the quality (the latter having a more general and a more specific designation (strength, and 60% strength, respectively). In addition there would be relations, e.g. to the relevant patient, and between quality and muscle group, as well as instantiations between these particulars and the corresponding universals.

## Conclusion

We have presented an example of the sorts of analysis needed to be performed when integrating an EHR application with the Referent Tracking paradigm. Central to this stage of the analysis is the decomposition of EHR data in such a way as to yield explicit reference to the particulars to which it refers. A variety of implicit references are uncovered by following the dependency relations between particulars as described in Basic Formal Ontology.

The analysis that we have provided, while abbreviated, contains an explanation of the methodology so that others may perform similar

efforts upon other EHR systems. These integration efforts will be rewarded by being the needed platform on which unambiguous communication between health care providers (and software agents) can be built. The analysis has made us even more aware of the importance of having a sound ontology such as BFO against which the decomposition of data can be performed.

## Acknowledgements

We would like to thank Medtuity Inc. for their graciousness in providing us with an installation of the MedtuityEMR application on which we conducted our research. In particular, we thank Matt Chase, CEO of Medtuity for his support on technical issues regarding the application.

## References

1. Ceusters W, Smith, B. Tracking Referents in Electronic Health Records. In: Engelbrecht R., Lovis, C., (eds) *Connecting Medical Informatics and Bio-Informatics: Proceedings of MIE 2005*. Amsterdam: IOS Press; 2005;71-76.
2. Ceusters W, Smith, B. Strategies for referent tracking in electronic health records. *Journal Biomedical Informatics*. 2006;39(3):362-378.
3. Manzoor S, Ceusters, W. Referent Tracking System. [cited July 9, 2007]; Available from: <http://sourceforge.net/projects/rtsystem>
4. Kent, W., The Entity Join. In: *Fifth International Conference on Very Large Data Bases*, (Rio de Janeiro, Brazil, 1979), Morgan Kaufmann Publishers, 232-238.
5. Grenon P, Smith, B., Goldberg, L. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In: Pisanelli DM, editor. *Ontologies in Medicine*. Amsterdam: IOS Press; 2004;20-38.
6. Ceusters W, Elkin, P., Smith, B. Referent Tracking: The problem of Negative Findings. In: Hasman A, editor. *Studies in Health Technology and Informatics*: IOS Press; 2006; 741-746.
7. Manzoor S, Ceusters, W., Rudnicki, R. Implementation of a Referent Tracking System. This volume. 2007.
8. 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):R46.
9. PATO\_Group. Phenotype and Trait Ontology. [cited March 9, 2007]; Available from: [http://www.bioontology.org/wiki/index.php/PATO:Main\\_Page](http://www.bioontology.org/wiki/index.php/PATO:Main_Page)