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

**Ron RUDNICKI [1] , Werner CEUSTERS[1,2] , Shahid MANZOOR[1], Barry SMITH[1,2]**
**[1] Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, NY, USA**
**[2] National Center for Biomedical Ontology, University at Buffalo, NY, USA**

**Abstract**

*The Referent Tracking paradigm which advocates the use of instance unique identifiers to refer to the entities comprising the subject matter of patient health records 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 to make it accord with the tenets of Referent Tracking. We describe the ontological principles on which this decomposition is based in order to allow integration efforts to be applied to other EHR applications by interested parties. We find that an ordinary statement from an EHR reveals a surprising amount of "hidden" data that is revealed by its decomposition according to these principles.*

## Introduction

The Referent Tracking (RT) paradigm was introduced in 2005[1] and its requirements and infrastructure were detailed in 2006.[2] The goal of the paradigm is to reduce ambiguous reference within the electronic health record (EHR) by introducing unique identifiers 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. RT identifiers are *globally* unique and their management is performed by a referent tracking system (RTS) designed to deliver services to EHR applications installed at separate locations in a health care network.[3] 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.

The problem of ambiguous reference in the EHR creates an obstacle 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 patient record is evidence of multiple separate instances of that condition or of multiple observations of the same instance.

Another significant benefit of RT is its capacity to do justice to the preservation of identity of an entity even as that entity changes from one type to another. The statements: "X has a dysplastic nevus at time t1" and "X has a malignant melanoma at time t2" 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 us (or software agents) to follow an entity as it evolves over time and this holds much promise for applications in domains such as post-marketing surveillance and the determination of patient outcomes. The same facility allows us to keep track of an entity as our knowledge evolves over time.

## Objectives

The 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 make it compatible with the requirements of RT, namely that the particulars assigned an IUI (for 'instance unique identifier') are instances of the kinds included in Basic Formal Ontology (BFO).[4] 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 primary care as well as 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.

One of the more remarkable features of MedtuityEMR is that it enables a user to generate a fully readable, highly detailed progress note using only point-and-click controls as input. The developers at Medtuity accomplished this 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 throughout the application to document the patient encounter. 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.

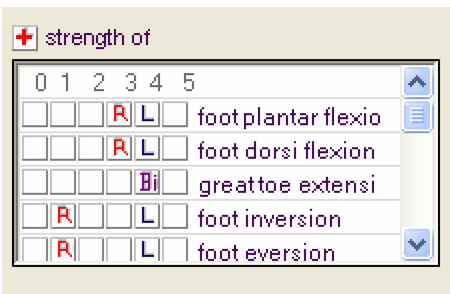


**Figure 1. MedtuityEMR '6-check' checklist control with measures of strength entered for flexions of a patient's feet.**

Since there is no qualitative difference between the data captured by the simple and more complex versions of a control type, choosing a single control upon which to illustrate our analysis is sufficient, we believe, to accomplish our educational goals. Our choice was a 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 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 dependency relations that hold between the various types of particulars as described in BFO, and that, as explained further down, lead to the distinction of three types of particulars relevant for our purposes: (1) independent continuants (e.g. John Smith's left femur), (2) dependent continuants (e.g. the fracture of John Smith's left femur), and (3) occurrents (e.g. the healing of the fracture of John Smith's left femur)

Data elements which refer directly to independent continuants require no expansion. Data elements which refer to one of the other types of particulars 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. For example, 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 Doe's left femur is an independent continuant – there is no other particular on which it depends in this ontological sense. The fracture of John Doe's left femur, in contrast, depends ontologically upon another continuant particular: John Doe's left femur.

Each of these distinctions among entities is mutually exclusive and pair-wise disjoint. Logically, they yield a total of 4 different categories of particulars. However, since all occurrent particulars are dependent entities (they all require one or more independent particulars which serve as their bearers) we are left with a total of 3 categories of particular entities: dependent and independent particulars on the one hand, and occurrents on the other.

## Referent Tracking

The first step in making an EHR application compatible with RT is to make an analysis of how data from the EHR application need to be restructured. To accomplish this we must complete, for each assertion in an EHR, the following tasks (based upon the distinctions amongst entities as described in BFO):

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,[5] and
6. identify the association of a customary name to a particular.

Furthermore, RT requires information about the state of affairs in reality to be expressed by means of one of the following types of statements:

1. the assignment of a IUI to a particular (e.g. that #321 stands proxy for John Doe and #7865 for John Doe'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. after the time #7865 broke and before the resulting pieces have grown back together),[5]
6. the association to a particular of a customary name (e.g. that #321's name is 'John Doe'; note that assigning an IUI to a particular is independent of whether or not that particular is assigned a name), and
7. the meta-description of a statement, that it has been added to the RTS.[6]

**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. While 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 a globally unique identifier which is formed through the concatenation of the identifier it assigns to the patient session during which the control is used, with the identifier it assigned to the control itself. The latter identifier is the same independent of for which patient or during what session it is used. However, the 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, or even prevent reasoning to occur 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[7]) 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 fool 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 can be done at various levels of detail, but for the purposes of the analysis, it is sufficient to do so 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)
- ObjectAggregate (independent continuant)
- Disposition (dependent 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** U1 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. This also involves checking whether an entity identified does not already have an assigned IUI.
2) For each entity determine its BFO category.
3) If an entity is an independent continuant, then no further expansion is required. 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.

Obviously, these steps need to be performed only once, i.e. 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 throughout if errors are to be avoided. One example: dispositional qualities like strength inhere only in continuants and not in occurrents. This guides the assignment of the patient's strength to his muscle group rather than to his act of flexion. 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 a more detailed analysis, 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,[8] this statement can be simplified into an entity-quality statement. Under this interpretation, there would be three particulars – the patient, the muscle group, and the quality – and two relations: patient having muscle group as part and quality inhering in muscle group, and instantiations between these particulars and the corresponding universals. This alternate treatment, suggested by the PATO ontology is offered with caution as this ontology does violate principles advocated by BFO. Namely, the PATO ontology assigns qualities to occurrents, which is in our view an example of treating the referent of a concept (an epistemological entity) as though it were an entity of the spatiotemporal world around us (an ontological entity). Thus, much further research is needed to determine if this alternate treatment of the data element is in fact compatible with the tenets of RT.

## Conclusion

We have presented an example of a portion of the 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 into referents of the particulars to which it refers both explicitly and implicitly. The 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 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. GA, Lovis, C., (eds) Connecting Medical Informatics and Bio-Informatics: Proceedings of MIE 2005. Amsterdam: IOS Press; 2005. p. 71-6.
2. Ceusters W, Smith, B. Strategies for referent tracking in electronic health records. Journal Biomedical Informatics. 2006;39(3):362-78.
3. Manzoor S, Ceusters, W. Referent Tracking System. [cited March 9, 2007]; Available from: http://sourceforge.net/projects/rtsystem
4. 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. p. 20-38.
5. 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. p. 741-6.
6. Manzoor S, Ceusters, W., Rudnicki, R. Implementation of a Referent Tracking System. Under Review. 2007.
7. 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.
8. PATO_Group. Phenotype and Trait Ontology. [cited March 9, 2007]; Available from: http://www.bioontology.org/wiki/index.php/PATO:Main_Page