ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

Representing Local Identifiers in a Referent-Tracking System

William R. Hogan[1], Swetha Garimalla[1], Shariq A. Tariq[1], Werner Ceusters[2]

[1]Division of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA
[2]Center of Excellence in Bioinformatics and Life Sciences, Buffalo, NY, USA

Keywords: realist ontology, referent tracking, electronic health records

1 Introduction

Referent Tracking (RT) is a novel, principled approach, suitable to a diversity of applications, to store data about particulars in reality. RT is unique in that it (1) assigns globally unique singular identifiers for *each* entity in reality about which information is stored, rather than for only some obvious entities such as, in the case of electronic health records (EHRs), the patient and his caregivers, and (2) uses a series of templates for unambiguous representation of the relationships of particulars [1-4].

At the University of Arkansas for Medical Sciences (UAMS), we are investigating RT's ability to handle diagnoses, procedures, demographics, encounters, hypersensitivity, and observations as they are reported in EHRs. The questions we are addressing are, amongst others, ones of representational adequacy: is it possible to represent in a referent tracking system (RTS) – with the current facilities the RT approach provides – the same entities in reality that EHR data are about?

A problem at this time, however, is that there are no such EHR systems. Existing EHRs – like numerous other biomedical software applications – use unique internal identifiers[1] to denote only those particulars that are persons, healthcare encounters, medical records, etc., but not diseases, injuries, disease courses, tumors, etc. For the sake of interoperability with non-RTS EHRs that cannot store mappings between local identifiers and IUIs, we require the ability to include local identifiers in the RTS. In this work, we confronted the issue of how to handle local identifiers in an RTS without violating the underlying ontological principles. The essence of the approach is to represent local identifiers just as we represent other entities external to the RTS: with instance unique identifiers (IUIs). Doing so however raised additional issues. The question we address here is whether we can solve these issues using existing RTS capabilities as embedded in the RT templates.

We use the following hypothetical scenario to illustrate our approach by building up a set of RT templates that represents the scenario: *Mrs. Smith is a new patient at ABC Medical Clinic, where an EHR has been in operation since 2005-05-05. On 2011-01-01, she checks in at the front desk and a member of the clinic staff enters her basic demographic and insurance information and creates a medical record for her in the EHR. After her visit with Dr. Jones, Mrs. Smith checks out and leaves the clinic.*

In our scenario, we assume (1) appropriate distinction between person identifiers and medical record numbers (MRNs), (2) the use of a single person identifier regardless of participation in an encounter as doctor or patient, (3) that the identifier 'EHR00001'[2] refers to the EHR instance at ABC Medical Clinic and is a local identifier in that EHR, and (4) that the local identifier 'O00001'uniquely denotes the organization. We ignore the building in which ABC Medical Clinic operates as it is not essential to our scenario, although this entity too requires a unique local identifier.

[1] We will from now on use the term 'local identifier' for the identifiers used in EHR systems to distinguish them clearly from the identifiers used in a referent tracking system.

[2] A note on use vs. mention: when we mention an identifier, we set it in single quotes. On the other hand, when we use an identifier to refer to something, we set it in italics.

2 The Approach

The approach is based on the fact that local identifiers and systems of such identifiers as used in healthcare organizations are as real and as external to the RTS as are the entities that they denote and can thus be denoted by IUIs too. In what follows, we use shorthand notation for IUIs, such as IUI_{Smith}, to improve readability.

We first assign IUIs to entities in the scenario using the RTS assignment or A-template: Mrs. Smith, Dr. Jones, ABC Medical Clinic, the EHR, and Mrs. Smith's medical record: $A<$ ***IUI$_{Smith}$,*** IUI_a, t_{ap} $>$, $A<$ ***IUI$_{Jones}$,*** IUI_a, t_{ap} $>$, $A<$ ***IUI$_{ABC}$,*** IUI_a,t_{ap} $>$, $A<$ ***IUI$_{EHR}$,*** IUI_a, t_{ap} $>$ $A<$ ***IUI$_{SmithRecord}$,*** IUI_a, t_{ap} $>$. The identifiers IUI_a and t_{ap} denote the entity that made the IUI assignment and the time of the assignment, respectively. See Ceusters [1] for a complete description of RT templates; here we provide enough explanation to illustrate our approach.

Each local identifier belongs to some system of identifiers. The system of identifiers is almost always constructed with the goal that each identifier in the system has one, unique, unambiguous reference. In our scenario, we included five identifier systems, because in our experience most EHRs have different systems for persons, encounters, etc. However, one system is feasible so long as all its identifiers uniquely denote one entity (for example, no encounter identifier would have the same string of characters as an MRN). What follows is the representation of the person identifier system in the EHR. We treat each of the other four systems (encounter, MRN, organization, and EHR identifier) in the same manner.

First, we assign an IUI to the person identifier system: $A<$ ***IUI$_{PersonIdSystem}$,*** IUI_a,t_{ap} $>$ and assert that it is an instance of a central identifier registry using a Particular-to-Universal (PtoU) template and the appropriate universal representation from the Information Artifact Ontology (IAO): PtoU$<$ IUI_a, t_a, ***inst, http://purl.obolibrary.org/iao, IUI$_{PersonIdSystem}$, IAO_0000579,*** $t_r$$>$. So that humans can differentiate among the various identifier systems represented in the RTS, we assign a name to each one using a Particular-to-Name (PtoN) template, which has the form: $PtoN<$ IUI_a, t_a, IUI_c, IUI_p, n, nt, t_r $>$ where IUI_c is the IUI for the entity that uses the name n, IUI_p is the IUI for the entity with the name n, nt is the name type (e.g., first name), and n the name associated with IUI_p. For the person identifier system, we use $PtoN<$ IUI_a, t_a, ***IUI$_{ABC}$, IUI$_{PersonIdSystem}$, 'internal system name', 'Person id system'***, $t_r$$>$. We similarly assign each local identifier an IUI, assert that it is an instance of centrally registered identifier, and assign it a name (the identifier string is the name n, e.g. 'O00001'). Each local identifier is part of its identifier system and denotes some entity, and each identifier system is part of the EHR: we represent these relationships for each identifier and system using Particular-to-Particular (PtoP) templates. The full set of templates for the scenario is publicly available online as a Google document: https://spreadsheets.google.com/ccc?hl=en&key=t8v6oS7tNl84OMDjuy5p2kQ&hl=en#gid=0

3 Discussion

We successfully represented EHR local identifiers in an RTS using existing RT facilities. Our approach is general, and could be used to represent local identifiers and identifier systems in any non-RTS.

The approach has certain advantages. Besides the already mentioned disambiguation of what exactly is denoted by a local identifier, distinct units within one organization can continue to use local identifiers despite referencing the same entities, and this without the need for complex identity-negotiation systems [7], or the need for an a priori agreement on a fixed set of entity types [8]. Thus it solves many issues the traditional federated database approach suffers from [9]. When EHRs of distinct organizations that provide healthcare to overlapping patient populations are connected to the same RTS or to RTSs which are connected in an RTS network [3], the approach enables tracking of the variety of identifiers used within these organizations. And when extended to include local dictionaries within units or organizations, the approach provides the additional benefit of implementing Smith's proposal to counteract the drawbacks of traditional controlled vocabularies and terminologies by using EHR data as a means to quality-control them (and

thus for purposes of automatically generating improved versions of such dictionaries) [10].

A drawback of the approach is that the PtoN-template, and more specifically the "name type" slot of that template, might become overloaded in its own right, and that at some point a name-type system might become necessary to track the various sorts of name types in use. Also, the approach leaves a number of relationships implicit, for example, that the systems of identifiers are endorsed by the organizations in whose EHRs local identifiers thereof are used. This problem could, in a naïve way, be solved by adding additional PtoP templates for which rather ad hoc relationships such as 'endorses' need to be defined. This sort of solution clashes however with the principles of Ontological Realism [11] to which RT aims to adhere. A better approach, and the topic of future work, is to introduce denotational bonds as proposed by Ceusters [12].

4 Conclusion

We identified a need to represent local identifiers and systems of such identifiers in EHRs in our work on RT. Prior to this work, whether and how RT could enable such a representation was an open question. The answer was affirmative: we successfully developed the required representations in an RTS and that made use of existing RT facilities. The approach nevertheless has some limitations we intend to address in future work by developing an ontological theory of denotational bonds.

References

1. Ceusters W. Dealing with mistakes in a referent tracking system. Ontology for the Intelligence Community (OIC-2007); Columbia, Maryland. 2007. p. 5-8.
2. Ceusters W, Elkin P, Smith B. Negative findings in electronic health records and biomedical ontologies: a realist approach. Int J Med Inform. 2007;76 Suppl 3:S326-33.
3. Ceusters W, Manzoor S. How to track absolutely everything. In: Obrst L, Janssen T, Ceusters W, editors. Ontologies and Semantic Technologies for Intelligence. Amsterdam: IOS Press; 2009.
4. Ceusters W, Smith B. Strategies for referent tracking in electronic health records. J Biomed Inform. 2006 Jun;39(3):362-78.
5. Rudnicki R, Ceusters W, Manzoor S, Smith B. What particulars are referred to in Electronic Health Record data? A case study in integrating Referent Tracking into an EHR application. AMIA Annu Symp Proc. 2007:630-4.
6. Ceusters W, Smith B. Referent tracking for digital rights management. International Journal of Metadata, Semantics and Ontologies. 2007;2(1):45 - 53.
7. Meghini C, Doerr M, Spyratos N. Managing co-reference knowledge for data integration. In: Kiyoki Y, Tokuda T, Jaakkola H, Chen X, Yoshida N, editors. Information Modelling and Knowledge Bases. Amsterdam, The Netherlands: IOS Press; 2009. p. 224-44.
8. Bouquet P, Stoermer H, Niederee C, Maa A. Entity Name System: The back-bone of an open and scalable web of data. Proceedings of the IEEE International Conference on Semantic Computing, ICSC 2008, number CSS-ICSC 2008-4-28-25: IEEE Computer Society; 2008. p. 554-61.
9. Heimbigner D, McLeod D. A federated architecture for information management. ACM Trans Inf Syst. 1985;3(3):253-78.
10. Smith B, Ceusters W. An ontology-based methodology for the migration of biomedical terminologies to electronic health records. AMIA Annu Symp Proc. 2005:704-8.
11. Smith B, Ceusters W. Ontological realism as a methodology for coordinated evolution of scientific ontologies. Applied Ontology. 2010;5(3-4):139-88.
12. Ceusters W. Introduction to Referent Tracking. Tutorial co-located with the International Conference on Biomedical Ontology (ICBO), Buffalo NY and part of the 2-day class From Basic Formal Ontology to the Information Artifact Ontology, 2009: Available from: http://www.bioontology.org/wiki/index.php/Ontology_Training:_Video_and_Audio_Presentations.