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, 4301 West Markham St., Little Rock, AR 72205-7199
{ WRHogan, SGarimalla, SATariq }@uams.edu

[2]Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Suite B2-160, Buffalo, NY 14203
ceusters@buffalo.edu

Abstract. First-order entities such as particular persons, healthcare encounters, organizations, and so forth, are denoted in electronic health records (EHRs)—in addition to other biomedical software applications—by means of unique identifiers that follow local format conventions. Such identifiers are second-order entities which in EHRs stand proxy for first-order entities. Referent-tracking systems (RTS), by contrast, require the use of instance unique identifiers (IUIs) to denote entities, and these IUIs must be globally unique and singular, and be used in formally defined RTS-templates. In our work on adapting EHR data to RTS, we found no explicitly defined procedure for handling such second-order local identifiers within an RTS. We developed an approach that represents local identifiers in the same way as an RTS represents other entities: by assigning them an IUI in their own right. This required us to introduce explicitly the 'denotes'-relation in the framework and to treat the IUI for a local identifier as if it were a name. An analysis of this approach uncovers its terminological nature and suggests a more formal, ontological approach that is the subject of future work.

Keywords: realist ontology, referent tracking, identifiers, electronic health records

1 Introduction

In a series of publications, Ceusters and colleagues have presented Referent Tracking (RT) as a novel, principled approach for a diversity of applications to store data about particulars in reality. [1-4]. These applications include electronic health records (EHRs) [4-5], World-Wide Web sites, intelligence systems (as in Central Intelligence Agency) [3], and digital rights management [6].

The novelty of the RT approach is in the use of (1) globally unique singular identifiers for *each* entity in reality about which information is stored rather than only for some obvious entities such as, in the case of EHRs, the patient and his caregivers, and (2) a series of templates for which the abstract syntax and corresponding semantics is given in Table 1.

Table 1: data-templates for a RTS

Tuple name	***Attributes***	Description
A-tuple	$< IUI_a, IUI_p, t_{ap} >$	Act of assignment of *IUIp* to a particular at time t_{ap} by the particular referred to by author IUI_a
D-tuple	$< IUI_d, IUI_A, t_d, E, C, S >$	A D-tuple is inserted (1) to resolve mistakes in RTS, and (2) whenever a new tuple other than a D-tuple is inserted in the RTS. The particular referred to by IUI_d registers the particular referred to by IUI_A (the IUI for the corresponding A-tuple) at time t_d. E is either the symbol 'I' (for insertion) or any of the error type symbols as defined in [1]. C is the reason for inserting the A-tuple. S is a list of IUIs denoting the tuples, if any, that replace the retired one.
PtoP-tuple	$<IUI_a, t_a, r, IUI_o, P, t_r>$	The particular referred to by IUI_a asserts at time t_a that the relationship *r* from ontology IUI_o obtains between the particulars referred to in the set of *IUIs P* at time t_r.
PtoU-tuple	$<IUI_a, t_a, inst, IUI_o, IUI_p, UUI, t_r>$	The particular referred to by author IUI_a asserts at time t_a that the particular referred to by IUI_p instantiates – by means of the *inst* relation defined in ontology IUI_o – the universal *UUI* at time t_r.
PtoC-tuple	$<IUI_a, t_a, IUI_c, IUI_p, CUI, t_r>$	The particular referred to by IUI_a asserts at time t_a that at time t_r concept code *CUI* from terminology system IUI_c is an accurate term for IUI_p
PtoU(-) -tuple	$<IUI_a, t_a, r, IUI_o, IUI_p, UUI, t_r>$	The particular referred to by IUI_a asserts at time t_a that the relation *r* of ontology IUI_o does not obtain at time t_r between the particular referred to by IUI_p and any of the instances of the universal denoted by *UUI* at time t_r.
PtoN-tuple	$< IUI_a, t_a, nt_j, n_i, IUI_p, t_r, IUI_c>$	The particular referred to by IUI_a asserts at time t_a that n_i is the name of the nametype nt_j used by IUI_c to denote the particular referred to by IUI_p at t_r.

At the University of Arkansas for Medical Sciences (UAMS), we are investigating RT as applied to anonymized EHR data. Whereas Rudniki et al. showed how RT could handle EHR data about assessments of strength of foot movements [5], and Ceusters touched on temperature measurements and open reductions of fractures [4], we are investigating its ability to handle additionally diagnoses, procedures, demographics, encounters, hypersensitivity, and observations as they are reported in EHRs. The questions we are addressing in our work on RT are, amongst others, questions 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? Once an EHR exists with adequate functionality and that uses nothing but RTS as a representational backend, we can then study the benefits/costs of RT as applied to EHRs.

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 particulars that are persons, healthcare encounters, medical records, organizations, physical locations (e.g., in buildings), etc. For the sake of interoperability with non-RTS EHRs that cannot store a mapping between local identifiers and IUIs, we require the ability to include these local identifiers in the RTS. In this work, we confronted the issue of how to handle local identifiers used by EHRs in an RTS without violating the underlying ontological principles.

Here, we describe collaborative work that the authors at UAMS did with author WC—one of the inventors of RT at the University at Buffalo—to address this issue. The essence of the approach is that we represent local identifiers just as we represent other entities external to the RTS: with instance unique identifiers (IUIs). Doing so however raises additional issues. The question we address here is whether existing RTS capabilities as embedded in the templates are adequate to solve them.

2 A hypothetical scenario

We use the following hypothetical scenario to illustrate our approach by building up a set of RT templates that represents the scenario. We do not show the metadata template associated with each RT template generated, because (1) the function of instantiated metadata templates is to track authorship of assertions, the evolution of what was known at each point in time, and the correction of mistakes [1] and none of these issues is being addressed in our scenario, (2) our solution neither relies on nor adds anything unique to metadata and (3) leaving out metadata improves clarity of exposition.

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 basic demographic and insurance information about her and creates a medical record for her in the EHR. After her visit with Dr. Jones, Mrs. Smith checks out and leaves the clinic.

The EHR creates and uses the following local identifiers to denote entities that participated in the clinic encounter:

[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.

Table 2: local identifiers used in the context of a medical encounter

Local Identifier	**Denotes**	**Identifier system**
P00094	Mrs. Smith	Person identifier
M00057	Mrs. Smith's record	Medical record number
E00124	Mrs. Smith's visit	Encounter identifier
P00012	Dr. Jones	Person identifier
O00001	ABC Medical Clinic	Organization identifier

In our experience, there are several issues with how EHRs typically use local identifiers. First, EHR users sometimes use one number to refer to both the person and that person's medical record. Thus the medical record number (MRN) is often overloaded in the sense that it denotes[2] both the person and her medical record. A common exception is when patient registration occurs in another system, which subsequently sends information about the registration (including MRN) to the EHR. Second, when a single person is both a patient and a clinician, EHRs often use different identifiers to refer to that person as a doctor and as a patient. Third, EHR information models are sometimes not designed carefully enough to distinguish between references to organizations and to the building(s) in which these organizations operate. Thus, some local identifiers are ambiguous as to whether they denote an organization or a building (or even a location inside a building). Fourth, and finally, EHR (and other software) installations may themselves have a unique local identifier. That is, if three organizations independently install an EHR from vendor *A*, there are three separate instances of *A*'s EHR. Sometimes organizations assign these EHR instances a unique identifier.[3]

In our scenario, we assume (1) appropriate distinction between person identifiers and MRNs, (2) the use of a single person identifier regardless of participation in an encounter as doctor or patient, (3) that the identifier 'EHR00001'[4] 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.

[2] We use the term 'denotes' to express the relationship between an identifier (or term in general) and the real-world referent for which the identifier is a proxy, and 'refers' to describe the portion of reality in which some agent uses a proxy to denote a real-world referent.

[3] Author WRH once worked for an organization that had three separate instances of one vendor's EHR and referred to them uniquely as 'H1', 'H2', and 'H3'.

[4] 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.

3 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. Which is to say, we also assign instance unique identifiers (IUIs) to these identifiers and identifier systems[5]. Note that we are dealing with the problem of local identifiers in EHRs that are not also RTSs. If an EHR were an RTS, all the aforementioned entities in our scenario would have IUIs and not "person ids", etc. The remaining issues then are how to relate local identifiers to various other entities and where in RT templates to store the character strings that make up local identifiers.

In what follows, we use shorthand notation for IUIs, such as IUI_{Smith}, to improve readability. In practice, RTSs use currently the Internet Engineering Task Force standard for Universally Unique Identifiers (UUIDs) for IUIs. For example, the string to reference the IUI for Mrs. Smith would be '0472c6a0-3de9-11e0-9207-0800200c9a66'.[6]

Before we begin, we first assign IUIs to various 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:

$$A< \boldsymbol{IUI_{Smith}}, IUI_a, t_{ap} >$$
$$A< \boldsymbol{IUI_{Jones}}, IUI_a, t_{ap} >$$
$$A< \boldsymbol{IUI_{ABC}}, IUI_a, t_{ap} >$$
$$A< \boldsymbol{IUI_{EHR}}, IUI_a, t_{ap} >$$
$$A< \boldsymbol{IUI_{SmithRecord}}, IUI_a, t_{ap} >$$

The identifiers IUI_a and t_{ap} denote the entity that made the IUI assignment and the time of the assignment, respectively. For reasons of clarity of exposition, we don't deal with them further.

3.1 Representing a system of identifiers

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,

---

[5] To be completely correct, such an IUI is in our approach, and in line with the Information Artifact Ontology, assigned to the generically dependent continuant which is exemplified by each specific occurrence of the local identifier and not to each such occurrence itself. This is justified by the facts that (1) for the EHR purposes covered here there is no need to distinguish between occurrences of the same local identifier that are represented in distinct colors, fonts, font sizes, and so forth, and (2) the principles of referent tracking state that IUIs are to be given only to entities that are relevant for the purpose at hand.

[6] The UUID itself is a sequence of 16 bytes.

etc. However, there could be only one system of identifiers so long as all its identifiers uniquely denote one entity (for example, no person 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:

$A< \boldsymbol{IUI_{PersonIdSystem}}, IUI_a, t_{ap} >$

Next, we assert that this identifier system is an instance of a central identifier registry using 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>[7]

Note that *inst* denotes the “instance-of” relation. For the IUI_o parameter of the Particular-to-Universal (PtoU) template, we use the Uniform Resource Identifier (URI) of the IAO, although we could assign an IUI to the IAO as well. For the *UUI* parameter, we used the identifier the IAO assigns to the universal *CRID registry*. Finally, for the t_r parameter, we would want to use an expression that indicates that the PtoU assertion has held true since the EHR instance was created, i.e., *since 2005-05-05*.

Finally, so that humans can differentiate among the various identifier systems represented in the RTS, we assign a name to the person identifier system 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 = the IUI for the entity that uses the name *n*
IUI_p = the IUI for the entity associated with the name *n*
nt = the name type (e.g., first name, last name)
n = the name associated with IUI_p

For the person identifier system, we use the following PtoN template:

[7] Similar to the A-template, PtoU, PtoN, PtoP, and PtoU- templates all have IUI_a and t_a parameters. These parameters refer to the author and the time the author asserted the relation to hold respectively, and can be distinct from the time and authorship of the instantiated template, which are specified in the corresponding metadata template. We leave them here simply as IUI_a, t_a throughout.

$$PtoN< IUI_a, t_a, IUI_{ABC}, IUI_{PersonIdSystem}, \textit{‘internal system name’}, \textit{‘Person id system’}, t_r >$$

For the *nt* parameter, we use ‘internal system name’ to distinguish it from other name types in use and thus to facilitate searching names. The *n* parameter should be set to whatever name the identifier system has in the EHR. In our example, we used ‘Person id system’. The IUI_c parameter references the user of the name, which here is ABC Medical Clinic (IUI_{ABC}).

3.2 Representing each identifier in the system

We similarly assign each local identifier an IUI, assert that it is an instance of centrally registered identifier (CRID) using the appropriate representation from the IAO, and assign it a name. For Mrs. Smith’s person identifier ‘P00094’ (we handle other identifiers in the same manner), we have the following:

$$A< \mathbf{IUI_{MrsSmithsPersonId}}, IUI_a, t_{ap} >$$
$$PtoU < IUI_a, t_a, inst, \textit{http://purl.obolibrary.org/iao}, IUI_{MrsSmithsPersonId}, IAO\_0000577, t_r >$$
$$PtoN< IUI_a, t_a, IUI_{ABC}, IUI_{MrsSmithsPersonId}, \textit{‘local identifier’}, \textit{‘P00094’}, t_r >$$

We place the identifier’s associated string of characters in the *n* parameter. The IUI_c parameter is IUI_{ABC}, because ABC Medical Clinic uses ‘P00094’ as an identifier. We set the *nt* parameter to ‘local identifier’, although the PtoU template already identifies it as a CRID. We could use ‘name’ instead for *nt* here, without detriment to representational adequacy, but ‘local identifier’ provides redundancy to facilitate searching and human understanding.

For the t_r parameter of both the PtoU and PtoN templates, we need the time that the EHR assigned ‘P00094’ to Mrs. Smith (2011-01-01). In our experience, EHRs typically do not capture this time. Thus in practice, we use the earliest date of the person’s first encounter in the EHR, as patient or physician.[8]

3.3 Representing the relationships of the local identifier and identifier system

Each local identifier is part of its identifier system, and denotes some entity. We represent these relationships for Mrs. Smith’s person identifier using Particular-to-Particular (PtoP) templates:

$$PtoP< IUI_a, t_a, part\_of, IUI_o, <IUI_{MrsSmithsPersonId}, IUI_{PersonIdSystem} >, t_r >$$
$$PtoP< IUI_a, t_a, denotes, IUI_o, < IUI_{MrsSmithsPersonId}, IUI_{Smith} >, t_r >$$

---

[8] Unless of course, the person has distinct patient and physician identifiers. Then, we assign to the patient id (physician id) the date of the first encounter as patient (physician).

The systems of identifiers in our example are all part of the EHR. For the person identifier system, we have:

$$PtoP< IUI_a, t_a, part\_of, IUI_o, <IUI_{PersonIdSystem}, IUI_{EHR} >, t_r >$$

The full set of templates is publicly available online as a Google document.[9]

4 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. Using RT templates, we represented (1) identifiers, (2) the identifier systems of which they are a part, (3) the entity who uses identifiers to denote entities in reality, and (4) names of identifier systems. We also captured an identifier's relationships to its identifier system and the entity it denotes. Likewise, we captured the part-whole relationship between the identifier system and the software application (EHR).

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]. This approach is different from the traditional federated database approach which tries to use middleware technology to unite disparate databases so that applications can draw from them, yet being unaware of their underlying differences. In this approach, identifiers are generated naively by concatenating identifiers for software systems with the identifiers in the systems, with no reference to the organizations that sanction the identifiers and their denotation of an entity in the world [9]. This approach suffers from the drawback that if two systems use the same identifier from a third party, they cannot appropriately detect that the concatenated identifiers denote the same entity. Furthermore, unique identifiers in this approach are said to reference objects within the system [9], and not external, real-world objects, and thus the approach also suffers from use-mention confusion. Equivalence between two "internal objects", one from system A and one from system B, is really equivalence of denotation in this approach, and is not explicitly represented in any manner.

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].

[9] https://spreadsheets.google.com/ccc?hl=en&key=t8v6oS7tNl84OMDjuy5p2kQ&hl=en#gid=0

We assumed for the sake of developing the approach that local identifiers in EHRs are unambiguous. However, as we pointed out, this assumption is often incorrect. Future work involves relaxing it, which likely involves associating the identifier with each entity and allowing users (software and human) to disambiguate among them when querying the identifier. Furthermore, for the physician/patient identifier problem, we could disambiguate them by assigning the identifiers instead to the unique patient and physician roles inhering in that person. A query on the patient identifier could then indicate it would like an instance of person (as opposed to an instance of a role inhering in a person) as output. Also, we hope that this work serves to motivate correct usage of identifiers in EHRs.

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]. A denotational bond is a social entity like a law, prenuptial agreement, or constitution, which is brought into existence by one or more parties that explicitly state, for such and such a purpose, and applicable under certain conditions, that A denotes B. Parties that accept a denotational bond can stop doing so when they see fit, and parties that did not participate in its creation can decide to subscribe to it later. All such events could be represented explicitly without violating the principles of Ontological Realism.

5 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.

6 References

1. Ceusters W. Dealing with mistakes in a referent tracking system. Ontology for the Intelligence Community (OIC-2007); Columbia, Maryland2007. 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 Dec;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.