Replacing EHR structured data with explicit representations

Jonathan Bona[1] and Werner Ceusters[2]
[1] Department of Oral Diagnostic Sciences, University at Buffalo, 355 Squire Hall, Buffalo, USA
[2] Department of Biomedical Informatics, University at Buffalo, 921 Main street, Buffalo, USA

1 INTRODUCTION

As part of a project to develop a roadmap for the creation of a multi-center fully identified patient data warehouse involving all State Universities of New York State (SUNY), we've examined patient records stored in an EHR database to 1) determine what its contents are intended to represent, and 2) develop ontologically sound models based on the principles of Ontological Realism and Referent Tracking (Ceusters, Chiun Yu Hsu, & Smith, 2014; Smith & Ceusters, 2010). The exploration of the EHR database is driven by identifying the structures that contain answers to questions that might be obtained with relative ease using the EHR system's user interface but that are difficult to find by working directly with the database, for example: '*what diagnoses have been made about which disorders a specific patient is suffering from; when were those diagnoses made and by whom; what entities are those diagnoses about?*'

This abstract presents issues with the data model currently used in the EHR database and an approach to address them.

2 CHARACTERIZING DATA AND ISSUES

The EHR database presents several obstacles to properly understanding its contents. The intended meaning of its data elements is not explicitly specified, but implicitly depends on connections to the user interface, other software that uses it, workflows, etc. Nevertheless, it's possible to determine some of this by examining tables, their elements and the connections between them.

For example, the tables named `Person` and `Problem` are linked to healthcare processes. `Problem` entries are organized under `Problem Headers`, where each header entry is supposed to correspond to a single thing (diagnosis, procedure) and `Problem` entries are spread out in time each under its header and correspond to updates to the record made during encounters (Weed, 1968). By focusing on patterns of diagnoses stored in these tables, we have identified several ways in which the data fail to represent. These include: multiple entries standing for the same entity; single entries that stand for more than one entity; entries that might represent either more than one entity of the same type separated in time, or a single entity that persists over time; entries wrongly marked as *resolved* or *errors*; and wrong or outdated active entries.

3 EXAMPLE

Figure 1 shows selected `Problem` entries and their `Problem Headers` for four patients. Within an example each row shows the problems under a single header for the patient. Columns represent days with entries for the patient. A filled oval indicates a problem entry on that day. This figure shows ordering of updates but not their spacing in time. A table of header descriptions appears on the next page.

In **example 1**, two headers are created initially: **e1ph1**

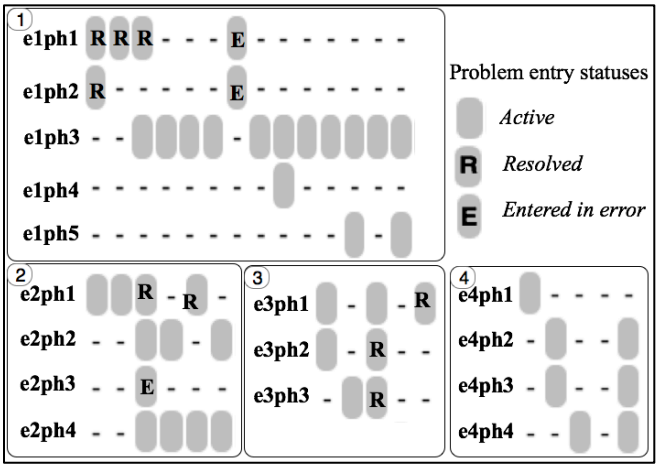

Figure 1 Examples of Problem entries for four patients

(*Diabetes mellitus type II*) and **e1ph2** (*Diabetes Mellitus With Complication*). Later **e1ph1** gets a new entry. Two months later, **e1ph3** (*Diabetes Mellitus*) is created. It is updated regularly. Six months after their creation, **e1ph1** and **e1ph2** are updated with new entries with the status *Error*. Later, **e1ph4** (*Type 1 Diabetes Mellitus - Uncontrolled*) and **e1ph5** (Diabetes *Mellitus With Complication*) are added. After that **e1ph3**, **e1ph4**, and **e1ph5** are updated occasionally, keeping the status *Active*.

The likely sequence of events is that the diagnosis of Type II DM was changed to Type I DM, after which a different header for Type I was created and used, and entries in the first two headers were marked *Error*. Old entries under those headers were marked as *Resolved*.

Clearly, this record has drifted from representing reality after just a few updates. This patient does not have more than one DM, but what was first thought to be Type I was then recognized as Type II. The system in this case fails to

distinguish 1) information that was retracted due to some change in knowledge about the patient's health from 2) information that was truly entered in error.

Example 2 shows a more correct use of the *Error* status: the header **e2ph3** (*Type II DM – Uncomplicated, Uncontrolled*) was created at the same time as headers about *ketoacidosis* (**e2ph2**) and *acanthosis nigricans* (**e2ph4**), which are complications of Diabetes. This mistake was quickly caught and **e2ph3** was marked *Error*. **e2ph1** here repeats the misuse of the *Resolved* status: the patient was thought to have Type I DM and the diagnosis changed in time. It's not that the Type I DM existed and then stopped existing.

Another issue is related to the use of ICD9 codes and ICD9-like descriptions: some Problem entries refer to multiple things. The patient's *Type II Diabetes* is one thing in the world; their *Ketoacidosis* is a separate related disorder.

A better representation would have identifiers for both the patient's diabetes and the patient's ketoacidosis, and would explicitly represent the relations between them (one was *caused by* or was a *complication of* the other).

e1ph1: Diabetes mellitus type II (NIDDM)
e1ph2: Diabetes Mellitus With Complication
e1ph3: Diabetes mellitus
e1ph4: DM (diabetes mellitus), type 1, uncontrolled
e1ph5: Diabetes mellitus with complication
e2ph1: Type 1 Diabetes Mellitus - Uncontrolled
e2ph2: Type II diabetes mellitus with ketoacidosis
e2ph3: Type 2 Diabetes Mellitus - Uncomplicated, Uncontrolled
e2ph4: Acanthosis nigricans
e3ph1: Closed Fracture Of The Shaft Of The Humerus
e3ph2: Closed Fracture Of Neck Of Femur - Transcervical
e3ph3: Closed Fracture Of The Humerus
e4ph1: Open Treatment Of Humeral Shaft Fracture w Plate/Screws
e4ph3: Fracture of left humerus

Example 3 shows a pattern of Problems for a patient who has multiple fractures. Entries in **e3ph1** and **e3ph2** indicate that the patient has closed fractures both of the shaft of the humerus, and of the neck of the femur. **e3ph3** (*Closed Fracture of the Humerus*) is created the following month. How many fractures are represented here: two or three? Probably the patient had two fractures: one in the femur and one in the humerus, and both **e3ph1** and **e3ph3** correspond to the same fracture. But possibly the patient had two fractures in the humerus at the same time, one specifically in the shaft and one in an unspecified place.

Example 4 shows a patient being treated for a humeral shaft fracture but there is initially no entry that represents the fracture itself – only for a treatment. **e4ph3** (*Fracture of left humerus*), which is created eighteen months after **e4ph1**, seems to represent the fracture itself. The next entry in **e4ph3** is *more than five years later*. Note that all Problems shown here have the status *Active*. None of these - including a six-year-old fracture - have been marked as *Resolved*.

4 EXPLICIT REPRESENTATIONS

We propose using ontology-based models that explicitly represent the entities and processes relevant to health care encounters. Figure 2 depicts an instance of *diagnosing a fracture* as part of a health care encounter.

Only some of the entities represented here will appear in other encounters. Any other diagnostic process will be a different instance. Its output will be a different instance than **diagnosis1** -- even if it's about the same things. The instance **diagnosis1** persists, even if it is later outdated, contradicted, or otherwise known to be wrong and marked as such. We also must be able to say that **disorder1** exists at time **t1** but not at time **t2**; that when it exists it is located at a particular spot on **bone1**; and that further fractures of **bone1** once **disorder1** has already healed are new fractures

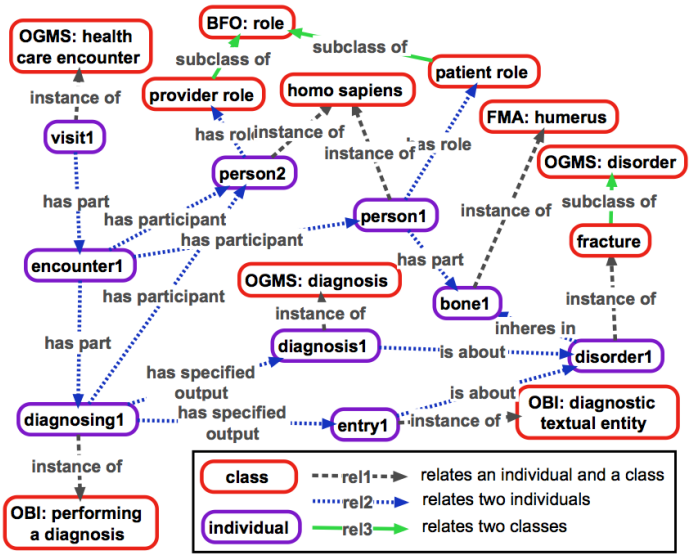

Figure 2 Representation of a fracture diagnosis

(**disorder2**). Having single identifiers for single entities (*this patient's diabetes*), which we can then say things about (*it is complicated by peripheral neuropathy*) is just as important as having separate terms for separate fractures.

Work is ongoing to develop computationally useful representations in OWL, mechanisms to interpret and translate patient data, and techniques to deal with temporal considerations and other issues that are not straightforward.

REFERENCES

Ceusters, W., Chiun Yu Hsu, & Smith, B. (2014). Clinical data wrangling using ontological realism and referent tracking. *CEUR Workshop Proceedings, 1237*, 27-32.

Smith, B., & Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology, 5*(3-4), 139-188. doi: Doi 10.3233/Ao-2010-0079

Weed, L. (1968). Medical records that guide and teach. *New England Journal of Medicine, 278*, 593-600.