Transforming Healthcare with the Internet of Things
J. Hofdijk et al. (Eds.)

doi:10.3233/978-1-61499-633-0-74

Ontological Foundations for Tracking Data Quality through the Internet of Things

Werner CEUSTERS[a,1] and Jonathan BONA[a]
[a] *Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA*

Abstract. Amongst the positive outcomes expected from the Internet of Things for Health are longitudinal patient records that are more complete and less erroneous by complementing manual data entry with automatic data feeds from sensors. Unfortunately, devices are fallible too. Quality control procedures such as inspection, testing and maintenance can prevent devices from producing errors. The additional approach envisioned here is to establish constant data quality monitoring through analytics procedures on patient data that exploit not only the ontological principles ascribed to patients and their bodily features, but also to observation and measurement processes in which devices and patients participate, including the, perhaps erroneous, representations that are generated. Using existing realism-based ontologies, we propose a set of categories that analytics procedures should be able to reason with and highlight the importance of unique identification of not only patients, caregivers and devices, but of everything involved in those measurements. This approach supports the thesis that the majority of what tends to be viewed as 'metadata' are actually data about first-order entities.

Keywords. Biological Ontologies, Internet, Metaphysics

Introduction

Although success stories for the use of electronic health records (EHR) to support individual patient care and biomedical research do exist , others argue that '*EHRs have yet to truly fulfil their promise to support clinicians in their patient care activities, including the essential work of building the patient's story*' . Also that the secondary use of EHR data to support, for instance, comparative effectiveness research is at least cumbersome because '*electronic health record data from clinical settings may be inaccurate, incomplete, transformed in ways that undermine their meaning, of unknown provenance, of insufficient granularity*' [4]. The major bottle neck for appropriate EHR use being the quality of data entry, this involves not only errors in human data entry but also the failure to enter data which are required [5].

It is precisely here that the Internet of Things (IoT) might bring tremendous advantages by avoiding the burden of structured data entry by humans through the connection of devices that use network services to enter data automatically. These devices, some of which not being designed specifically for healthcare purposes, will range from context-aware thermometers, weighing scales and tonometers to intelligent video and sound recorders that during a patient encounter – and if the patient so desires also during his everyday activities and insofar relevant to his health – record every

[1] Corresponding Author.

single event or state. Powerful analytics software will then have the capacity to extract all and only meaningful data from these recordings.

But as the use of EHR systems might itself constitute a risk for patient safety, so may the IoT lead to adverse events due to device malfunctioning or communication errors leading to erroneous data entry. However, whereas generally the odds for system malfunction increase relative to the number of devices that are part of the system, the IoT can be set up in such a way that these odds decrease by exploiting the fact that devices can observe and measure not only what is the case for the patient, but also for the patient's environment including the interconnected devices themselves! This requires an IoT for health not only to manage data about the patient but also about its own components and how these components contribute to assertions about the patient.

In this paper, we propose Ontological Realism as a methodology to identify and describe (1) which components within the ontological structure exhibited by the configurations of entities observed and measured by IoT devices are essential and (2) the abstract syntax towards which the output of IoT devices (or the subsequent interpretation thereof) should be formatted, for such devices and their operation to minimize both the burden of data entry and the risks for assertion errors.

1. Methods

Ontological Realism (OR) is a theory that defines the principles for high quality ontology development used in the Basic Formal Ontology (BFO) and the ontologies accepted in the Open Biomedical Ontology (OBO) Foundry [6]. Crucial for the proposal advanced here is that OR recognizes two major types of components out of which reality is built: (1) *particulars* such as this paper and its authors – all entities that carry identity, and (2) *universals*, for example those generic entities denoted by general terms such as 'paper', and 'person', which have particulars as instances. Particulars may enjoy *relations* with other particulars so as to form *configurations*. An example is the configuration which constitutes the ground truth for the assertion that this particular paper was the output of a particular collaborative writing process in which the particulars Werner Ceusters and Jonathan Bona, as well as their beliefs about the adequacy of the proposal advanced, all participated during certain time periods.

Referent Tracking (RT) is an OR-based paradigm for knowledge management that originally has been introduced in the context of EHR keeping [7]. Whereas realism-based ontologies focus on the *types* of particulars that exist in reality, RT focusses on the particulars themselves, more concretely on how assertions about the configurations formed between particulars and/or universals should be construed to maximally mimic the structure of reality. Key in RT is (1) the assignment – or reuse in case of former assignment – of instance unique identifiers (IUI) to *every* entity about which some assertion is made, and (2) the use of these identifiers in relational expressions following assertion templates that are maximally self-explanatory and unambiguous [8].

We demonstrate (1) how existing OBO-Foundry ontologies can serve as a source for the representation of all high-level entity types relevant to quality monitoring of devices and data analytic components connected in the IoT for Health, and (2) how RT is expressive enough to represent the relationships enjoyed by instances of these types, and can serve as a basis for analytics regarding the ground truth of assertions.

2. Results

Table 1 summarizes the types essential for managing data and metadata to be generated over the IoT for Health with data quality control in mind, specifically the quality aspects *accuracy, consistency and reliability*. Types are universals (U) as introduced in section 1, or defined classes (DC) grouping particulars on the basis of fiat demarcations relevant to some purpose, e.g. to distinguish patients from caregivers [6]. Types are elucidated (E) when primitive or defined (D) in terms of the necessary and sufficient conditions for instantiation. They are taken from BFO [6], the Ontology for General Medical Science (OGMS, [9]), ReMINE's adverse event ontology [10], and the Ontology of Biomedical Investigations (OBI, [11]) or introduced as subtypes from existing types. Further subtyping is possible, but is not relevant for our purposes here.

Table 1. Universals (U) and Defined Classes (DC) assessed essential for reporting and analyzing data and metadata generated over the IoT for Health. Terms used in a strict technical sense are formatted in small caps and are described either elsewhere in this table (printed in bold) or in the cited reference.

Type		Definition (D) or Elucidation (E)
Assay	U	(E) planned process to produce information about a material entity by physically examining it or its proxies [11]
Bodily feature	DC	(D) bodily component, bodily quality, or bodily process. [9]
Caregiver	DC	(D) Human Being in which there inheres a Caregiver Role
Device	U	(E) Object which manifests causal unity via engineered assembly of components & of a type instances of which are maximal relative to this criterion of causal unity. [6]
Interpretive process	U	(D) cognitive process (in brains or through software implementations) which brings into being, sustains or destroys cognitive representations on the basis of an **observation** [10]
IoT for Health	DC	(D) Object aggregate which is part of the IoT and is composed out of **Devices** and other Objects that generate or analyze **Observations** within a community of **Subjects of Care**.
Sensor Device	DC	(D) **Device** in which inheres the Functions to perform **Assays** and to generate **Observations**
Site	U	(E) 3-dimensional Immaterial Entity that is bounded by a material entity or is a 3-dimensional immaterial part thereof. [6]
Subject of care	DC	(D) Human Being undergoing Acts of Care [10]
Observation	DC	(D) **Representation** resulting from an assay [10]
Representation	DC	(D) Quality which is about or is intended to be about a Portion of Reality [12]

Table 2 lists just a few RT statements describing part of a portion of reality evolving over a temporal period *t* during which an inpatient (IUI #1), born at time *t1*, staying in the hospital wearing an RFID tag (#2) since *t2*, is clinically examined in an exam room (#3). The room has an RFID sensor (#4) which is connected to the hospital's IoT for health (#5) and which generated a representation (#109) of the location of the patient's tag when #1 entered room #3 at *t3*. A nurse (#6) measures (#7) at *t4* the patient's temperature (#8) with her personal digital thermometer (#9) which is also connected to #5 since *t5*. has a fingerprint reader to identify patients and a built-in RFID tag to locate its position in the building so that when at *t6* the nurse entered room #3 with the thermometer, sensor #4 generated a representation of its location (#117). When the patient touches the fingerprint reader of the thermometer, a picture (#10) of the patient's fingerprint pattern (#11) is transmitted at *t7* to a fingerprint analyser. This analyser determines (#12) on the basis of another picture of #11 already on file that #10 is about #11, as a result of which it sends the patient's IUI, i.e. '#1', to the thermometer. After the thermometer has registered a value for #10 at *t8*, a representation (#130) is generated asserting that the patient has a temperature of 37°C.

Table 2. Some RT statements, preceded by their own IUI, representing part of a scenario of taking a patient's temperature in a healthcare facility with an IoT for Health. The TYPES are listed in Table 1, the relations are defined in (or definable from) the corresponding ontologies, and the temporal operators (e.g. since, includes, during, …) defined in European Norm 12381: Time Standards for Healthcare Specific Problems [13].

#100:	#1	***instanceOf***	Human Being	***since*** t1	#101:	#1	***instanceOf***	SubjectOfCare	***since*** t2
#102:	#2	***instanceOf***	Device	***includes*** t2	#103:	#2	***locatedOn***	#1	***since*** t2
#104:	#4	***instanceOf***	Sensor Device	***includes*** t	#105:	#4	***locatedIn***	#3	***includes*** t
#106:	#3	***instanceOf***	Site	***includes*** t	#107:	#4	**partOf**	#5	***includes*** t
#108:	#5	***instanceOf***	IoT For Health	***includes*** t	#109:	#2	***locatedIn***	#3	***since*** t3
#110:	#4	***authorOf***	#109	***at*** t3	#111:	#6	***instanceOf***	Caregiver	***includes*** t
#112:	#7	***instanceOf***	Assay		#113:	#1..	***specifiedInputOf***	#7	***during*** t4
#114:	#6	***participantOf***	#7	***during*** t4	#115:	#9	***participantOf***	#7	***during*** t4
#116:	#9	***instanceOf***	Sensor Device	***includes*** t4	#117:	#9	***locatedIn***	#3	***at*** t6
#118:	#9	***partOf***	#5	***since*** t5	#119:	#9	***locatedOn***	#6	***includes*** t4
#120:	#4	***authorOf***	#117	***at*** t6	#121:	#11	***instanceOf***	BodilyQuality	***since*** t1
#122:	#11..	***inheresIn***	#1	***since*** t1	#123:	#10	***isAbout***	#11	***since*** t4
#124:	#12	***instanceOf***	Assay		#125:	#123	***SpecifiedOutputOf***	#12	***since*** t4
#126:	#11	***specifiedInputOf***	#12	***during*** t2	#127:	#8	***inheresIn***	#1	***since*** t1
#128:	#8	***instanceOf***	Bodily Quality	***since*** t1	#129:	#9	***authorOf***	#130	***since*** t4
#130:	#8	***instanceOf***	37°C	***at*** t8	#130:	#10	***instanceOf***	Observation	***since*** t7

3. Discussion

Representations of the sort exhibited in Tables 1 and 2, covering the totality of devices available within an IoT for Health rather than just in the partial scenario developed here, offer ample explicit information to feed algorithms for data quality monitoring by exploiting two specific features of the data collection methodology.

The first one is the multitude of sensor devices that can be used to monitor individual particulars from different perspectives. In the scenario sketched, it is both the RFID tag (#2) of the patient and the fingerprint reader in the thermometer (#9) that provide enough evidence to conclude that it is indeed patient #1 who is examined in room #3. Assertions #103 and #109 together, in combination with the axioms of the ontologies from which the ***locatedIn*** and ***locatedOn*** relationships are taken, provide an argument that #1 is in the room, although it might be the case that after #103 was asserted by the reception clerk who gave the RFID tag to patient #1, the patient lost it and picked up another one, or that the clerk made a typo. Similarly, assertions #117 together with #122 through #126 provide evidence for #1 being in the room. But if either something went wrong with the fingerprint analysis or with the RFID tag, both collections of assertions would not lead to the same conclusion what would be a trigger for further verification. The second feature, not worked out in detail in Table 2, is that the patient data can be used to monitor the proper functioning of the IoT devices. If the same scenario applied to several patients would lead to inconsistencies, then it is very likely that either sensor #4 or the thermometer are malfunctioning. This feature makes it clear that what is typically considered metadata, are actually data in their own right.

Although there is no shortage on papers that discuss security and confidentiality risks associated with the IoT for Health, the issue of data quality and anomaly detection is more scarcely dealt with as witnessed by a recent review [14]. An exception is [15] in which a mathematical model towards the reliability of sensor data is proposed, however it's only applicable to continuous sensors with high refresh rate measuring characteristics of ongoing processes (heart beats, continuous blood pressure monitoring) rather than between discrete events. Several papers discuss the potential use of ontologies in the IoT for Health, but here also mainly for security, e.g. [16]. A literature review over biomedical research papers published between 2001 and 2011 revealed an increasing amount of work on ontology, but little on ontological approaches to data quality [17]. The approach has two limitations. One is the use of ontological realism which is reported to be hard to understand [18]. The other one is the development of not only efficient, but also useful reasoners. Whereas the former requires more education, the latter is a matter of further research and development, including the design of an action logic for inconsistency detection and alerting.

Acknowledgement

This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health.

References

[1] Lugovkina, T. and B. Richards, Clinical events classification for using the EHR to provide better patient care. Stud Health Technol Inform, 2010. 156: p. 167-70.

[2] El Fadly, A., et al., The REUSE project: EHR as single datasource for biomedical research. Stud Health Technol Inform, 2010. 160(Pt 2): p. 1324-8.

[3] Varpio, L., et al., The EHR and building the patient's story: A qualitative investigation of how EHR use obstructs a vital clinical activity. Int J Med Inform, 2015.

[4] Hersh, W.R., et al., Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care, 2013. 51(8 Suppl 3): p. S30-7.

[5] Sparnon, E. and W.M. Marella, The Role of the Electronic Health Record in Patient Safety Events. Pennsylvania Patient Safety Advisory, 2012. 9(4): p. 113-121.

[6] Arp, R., B. Smith, and A.D. Spear, Building ontologies with basic formal ontology. 2015, The MIT Press,: Cambridge, Massachusetts. p. 1 online resource.

[7] Ceusters, W. and B. Smith, Referent Tracking in Electronic Healthcare Records, in Connecting Medical Informatics and Bio-Informatics. Medical Informatics Europe 2005, R. Engelbrecht, et al., Editors. 2005, IOS Press: Amsterdam. p. 71-76.

[8] Ceusters, W., Chiun Yu Hsu, and B. Smith, Clinical Data Wrangling using Ontological Realism and Referent Tracking. CEUR Workshop Proceedings, 2014. 1237: p. 27-32.

[9] Scheuermann, R.H., W. Ceusters, and B. Smith, Toward an ontological treatment of disease and diagnosis. Summit on Translat Bioinforma, 2009. 2009: p. 116-20.

[10] Ceusters, W., et al., An Evolutionary Approach to Realism-based Adverse Event Representations. Methods of Information in Medicine, 2011. 50(1): p. 62-73.

[11] Jensen, M., et al., Applications of OBI 'assay', in International Conference on Biomedical Ontology. 2014: Houston, TX. p. 96-97.

[12] Smith, B. and W. Ceusters, Aboutness: Towards Foundations for the Information Artifact Ontology, in International Conference on Biomedical Ontology. 2015: Lisbon, Portugal. p. 47-51.

[13] Ceusters, W., et al., TSMI: a CEN/TC251 standard for time specific problems in healthcare informatics and telematics. International Journal of Medical Informatics, 1997. 46(2): p. 87-101.

[14] Rassam, M.A., A. Zainal, and M.A. Maarof, Advancements of data anomaly detection research in wireless sensor networks: a survey and open issues. Sensors (Basel), 2013. 13(8): p. 10087-122.

[15] Marie, P., et al., From Ambient Sensing to IoT-based Context Computing: An Open Framework for End to End QoC Management. Sensors (Basel), 2015. 15(6): p. 14180-206.