An Ontology-based Model of Clinical Information

Thomas Beale[†a], Sam Heard MD[‡ab]

[†] CTO, [‡] CEO Ocean Informatics Pty, Ltd
[a] Visiting Senior Research Fellow, University College London
[b] Adjunct Professor, University Central Queensland

Abstract

In this paper we describe a model of clinical information designed to make health information systems properly interoperable and safely computable. The model is a response to a number of categories of requirements, ranging from the semantic to the performance of software at runtime. We argue that the starting point of a successful model must be an ontological analysis of the process of clinical care delivery, seen as a scientific problem-solving process. From this approach we develop a classification of types of clinical information called the Clinical Investigator Record (CIR) ontology.

Keywords:

EHR, health records, clinical process, information models, information systems, archetypes, problem-solving.

Introduction

Interoperability and computability of information in the healthcare sector promise a great deal in the global effort for better quality and more efficient healthcare, yet remain largely unachieved. Models of clinical information, including models of the Electronic Health Record (EHR) and variants (EMR, EPR, CPR etc), do not currently have a theoretical basis sufficiently strong to guarantee interoperability and computability. Any information model developed must also satisfy a myriad of other requirements when actually implemented including: computational efficiency and performance; economically viable implementation; maintainability; system scalability and extensibility; and the considerable privacy and security requirements of the health domain. A successful model of health information is thus like any other good model; it is an expression not only of the semantics of the domain, but also a response to the needs of economics of system construction and runtime performance.

Methods

Our methodological approach to developing a model of health information consisted of four steps. Firstly a conceptual model for understanding the creation and recording of information during the healthcare process was developed. This was used to develop an ontology of recorded health information. An information model suit-able for software development was developed based on the process model and ontology as well as other requirements - this became the *open*EHR Entry model. Finally we validated the model in a number of ways, including testing it with real clinical statements and queries and showing how it works in implemented systems. This paper describes the first two steps.

Background

Systems and models whose design ignores the real world in which they operate will fail to work well (see e.g. [1], [2]). The challenge is to find a way of basing a system design on the reality in which it is intended to function. An ontological approach provides a formal basis. An 'ontology' can be understood as a 'description of reality' and may be subdivided into a description of *process* (e.g what occurs, what is done) and a description of *information* (e.g. what is thought, said, communicated, remembered).

Sowa [3] provides a top-level and very general classification of the world, which is a useful reference point for any ontology, while more health-specific ontologies include the Ontology of Biomedical Reality (OBR) [4] and the SNOMED-CT clinical terminology [5]. The CIR ontology describes information created in the process of healthcare delivery.

A model of clinical information that will perform at all levels must meet three broad categories of requirement:

- *semantic requirements*: accurately represent and convey intended meaning of its users;
- *functional requirements*: provide the functions required by its users when deployed; and
- *economic requirements*: enable economically viable software construction and maintenance into the future.

A process model for clinical healthcare

It is important that any model used as a basis for software be based on a generic and high-level conceptualisation of the care delivery process. Various attempts to do this have been made in the past. Weed's problem-oriented medical record (POMR) methodology [6] formally linked a particular model of the process of care (relying heavily on testing) to the information gathered during that care (leading to the well-known "SOAP" headings). Elstein's

'hypothetico-deductive' model of clinical reasoning [7] mainly accounts for the cognitive aspects of clinical care (cue recognition and evaluation) during diagnosis. The model makes a good case for clinical healthcare to be seen as an iterative, scientific problem-solving process. The model originally developed by Rector, Nolan and Kay in the PEN&PAD system [8] is not so much based on problem-solving, rather a faithful yet flexible ability to *record* clinical action, thinking and dialogue. They propose only a limited ontology of information, namely 'direct observations' and 'meta-observations', thus distinguishing 'facts' from 'opinions'.

The Danish 'general electronic patient journal' (G-EPJ) [9], shown in Figure 1, includes a conceptual model of the iterative problem-solving process and categories of information generated. This model represents the clinical investigative process in the form of the cycle with process steps (circles) and information arising (blocks). While it has proved too rigid in practice, it formalises both process and information based on a paradigm of rational problem-solving.

Various clinical modelling efforts dating from the RICHE [10] project to the present HL7 version 3 standard [11] have based their models on an 'act management' paradigm in which all aspects of clinical care are represented as Acts. This approach enables 'everything that is done' to be recorded, which has clear attractions for business process tracking and cost-recovery (based on costs associated with fine-grained acts). It also has applications in messaging, where most notifications are to do with acts being requested, or being carried out.

Useful principles

A proper understanding of the notion of 'recording' is required. The point of view taken here is that what is recorded in the health record is likely to be a small, selective choice of notes about real events, situations etc, intended for interpretation by other professionals. The implication is that any model of health information must capture the *cognitive* communication processes of health professionals wherein very partial information may be recorded, rather than some more general notion of 'comprehensive fact representation'.
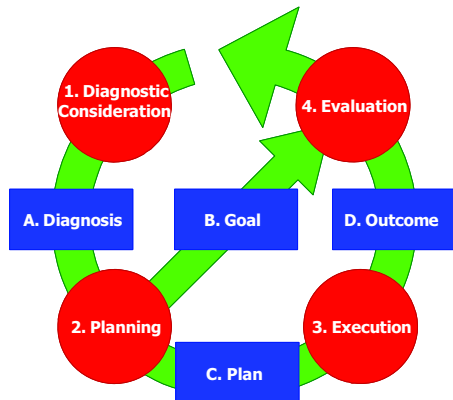

Figure 1 Danish G-EPJ Conceptual Model

It is also important to avoid the danger of confusion of linguistic entities with real entities and phenomena. This often manifests in models that define types with names like 'problem', 'observation', 'plan' and 'result', without properly defining the meaning of these categories.

The clinical investigator model

The theory we describe in this paper was originally conceived during the GeHR (Australia) [13] (pp29-32) and *open*EHR [14] projects, and built on some of the earlier work described above. We model health care delivery as two kinds of process: a *clinical process,* corresponding to the interaction between a 'clinical investigator system' and a 'patient system', situated within a *business process*, which is owned by an 'administrative context'. The model can be illustrated in two equivalent ways, as shown in Figure 2.
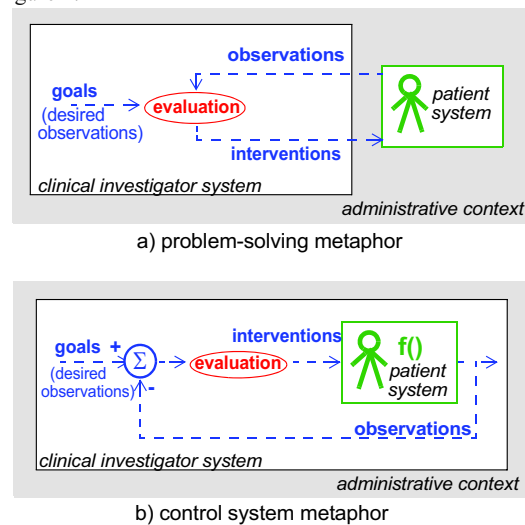

a) problem-solving metaphor

b) control system metaphor

Figure 2 - The Clinical Investigator Model

In part a) of this figure, the clinical process is shown as a two-entity system: a subject of investigation - the 'patient system', and the investigating entity - the 'investigator system'. The former is defined as the subject of care (typically one person, but may be more) seen essentially as a biological or social system (depending on the perspective of the investigator), while the latter is defined as the totality of healthcare professionals and other agents who perform actions to do with care of the patient; this includes the patient in the role of self-carer or self-medicator, as well as any family members or other people. The investigator system's purpose is to use *observations* as a basis for *evaluation*, leading to *interventions*. The investigator system is driven by *goals*.

Part b) of the diagram shows exactly the same elements, but redrawn in a manner familiar to control system engineers and those familiar with systems theory. In this schema, goals (desired observations) are seen as the controlling input signal, with observations being the output, and the patient the filter, or 'transfer function'. The difference between the actual and desired observations is used

by the controller (the evaluation activity) to determine the actual (corrective) input signal to the patient, in the form of interventions. In acute care, the external control system realised by the clinical investigator(s) is attempting to perform the same kind of homeostatic function that the body itself does, and in some cases, it is performing *exactly* the same function, such as in the case of patients using substances such as insulin. The fact that the clinical process can be mapped directly onto the standard model of negative feedback is not only intellectually satisfying, but indicative that this particular conceptualisation of the process is likely to be correct.

An ontology of clinical information

Since our main goal here is to understand how to support healthcare delivery using information systems, we initially need to know what kinds of information might be created by the processes described above *for recording and use in the information system*. We can redraw the investigator system in order to more clearly show the types of information created during the care process, as shown in Figure 3. Five types of information are identified, as follows:

- *observation*: information created by an act of observation, measurement, questioning, or testing of the patient or related substance (tissue, urine etc.), including by the patient himself (e.g. taking own blood glucose measurement), in short, *the entire stream of information captured by the investigator, used to characterise the patient system*;

- *opinion*: inferences of the investigator using the personal and published knowledge base about what the observations mean, and what to do about them; includes all diagnoses, assessments, plans, goals;

- *instruction*: opinion-based instructions sufficiently detailed so as to be directly executable by investigator agents (people or machines), in order to effect a desired intervention (including obtaining a sample for further investigation, as in a biopsy);

- *action*: a record of intervention actions that have occurred, due to instructions or otherwise;

- *administrative event*: a record of a business event occurring within the administrative context, such as admission, booking, referral, discharge etc.
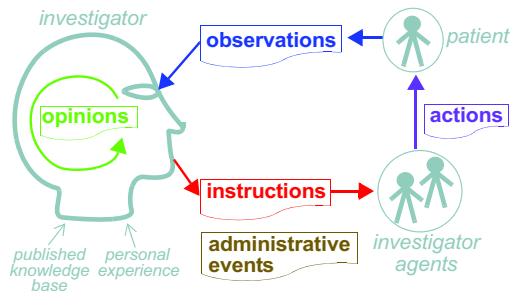

Figure 3 - Information Created by the Investigator

From these categories we construct an initial ontology, shown in Figure 4, the main purpose of which is to introduce the abstract categories 'recorded information' and

'care information', and to situate the five information types with respect to these.
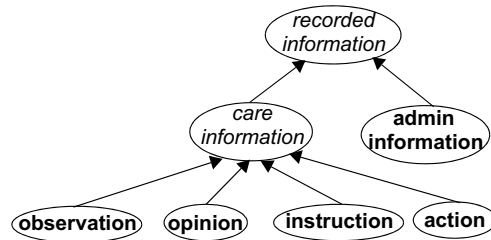

Figure 4 - Initial Clinical Information Ontology

Historical information: observation and action

We have already defined 'observations' as any information received by the clinical investigator used to characterise the patient state. It is axiomatic that observation information is 'in the past', since it expresses states or events that have already occurred. Observations are therefore *historical* in nature.

The Action information category is also historical in nature: a record of actions taken cannot be written until the actions have been completed (the documentation of actions 'to be done' is another category we call Instruction). Clinical actions such as administration of a medication or discontinuing a medication can all create Action instances. We can conclude, therefore, that the Observation and Action categories are subsumed by a category which we will simply call Historical (following Sowa's top-level categories). These categories can now be understood respectively as: a historical record of natural phenomena to do with the patient, and a historical record of things done to or for the patient.

Information in the cognitive 'present': opinions

Within the cognitive investigation process, clinical thinking, including any analysis, assessment or planning, is always 'in the present'. This is true of any analytical or decision-making activity: temporally it comes between evidence about the subject of study and actions to be performed in the future. When specific examples of clinical opinion are examined, the relationship with time may not always seem so clear. Consider a diagnosis; to properly characterise the problem, it will include time-related information such as the following:

- date of initial onset
- date clinically recognised
- for each occurrence or exacerbation:
  - date of last onset of occurrence
  - date of resolution
- date of resolution

Timings like this represent the 'continuant' or persistent relevance of this information and as such the Opinion category includes the idea of summary or review. As such the diagnosis is not itself an *observational* record of onset, exacerbation, resolution etc. (these events may have been documented as Observations when they occurred), but a

structured *summary* of dates and other items seen as important by the physician. The Opinion category of information is thoughts and analysis *about* previously recorded facts, not a historical record of the facts themselves. This is the same as items reported in the "news" section of a newspaper, and later analytical pieces, such as editorials and current affairs TV shows - the latter cannot help mentioning facts and dates from the former, but is not itself the contemporary record of the events in question. The Opinion category corresponds to Sowa's Description category.

Information about the future: instructions

We denote the last major category of information coming from the clinical care process as Instruction. Information in this category is in the form of orders, requests, and other directions to be performed by clinical investigator agents, so as to effect desired changes in the patient system, or else to gather new evidence for further consideration. Instruction time is therefore in the future, since the information speaks of events that have not yet occurred or are not completed. The Instruction category corresponds to Sowa's Script category.

We can describe two sub-categories of Instruction, distinguished on the basis of being for the purpose of investigation (to make further observations) or for intervention (aimed at changing the state of the patient system). We will call these categories Investigation-request and Intervention-request. As with the other categories, these names are to be understood as contractions for 'record of instruction for investigation' and 'record of instruction for intervention'. These may overlap when investigations have some treatment aspect; for example, removal of fluid around the lung will aid in the diagnosis as well as alleviate symptoms.

One of the key clarifications provided in the types described so far is between Instruction and Action. The former is a specification of what to do, while the latter is a record of what was done. A medication order may say to take 1 or 2 tablets every 4 hours, whereas an administration action will record the actual number taken. Models that are not clear about this distinction are likely to be problematic when connected to workflow systems. In simple terms, a series of Actions always represents a particular path taken through the set of possibilities described by an Instruction.

Accordingly, we arrive at an improved form of the ontology in which the three categories History, Opinion and Instruction are distinguished on the basis of the temporal relationship between the recorded information and the investigation process.

Observation and intervention sub-categories

We can make further improvements by analysing the categories so far in terms of their relationship to the two sides of the investigation process, i.e. 'observation' (characterising the patient system) and 'intervention' (changing it). We have included two subclassifications on this basis, namely Observation/Action and History and Investigation-request with Intervention-request below Instruction level.

Types of opinion

The top-level category *opinion* corresponds to inferences about observations, as well as about interventions, such as goals, and plans. Types of opinion have historically posed the most difficulties for models of health information, because they are so variable. The Opinion category corresponds to the notion of 'hypothesis' in general science. In health informatics, it corresponds to Weed's 'assessment' and (partially) 'plan' categories, to Van de Velde & Degoulet's [12] 'abstraction' category, and to Rector *et al's* [8] 'meta-observations'.

Here we propose two sub-categories, more directly related to the systems view of healthcare proposed above: *assessments*, i.e. opinions about what is happening in the Patient System, and *proposals*, i.e. opinions about what should be done by the Investigator system. Under this approach, we can say that the first category relates to *current* or *projected* states of affairs, and the second to a *desired* state of affairs.

The primary example of an Assessment in clinical medicine is the diagnosis. A diagnosis is the attachment of a label to a group of observed signs and symptoms, which designates it in the understanding of the Investigator as being a particular known phenomenon. A differential diagnosis, another category of assessment, allows for multiple possibilities, due to the lack of sufficient information or understanding to discount all but one.

Assessments include *quantification of risk* providing the basis for prevention or investigation. Such a notion includes the idea of *family history* being a risk arising from an unusual prevalence of a condition amongst relatives. *Prognosis*, the likely outcome corresponding to a current diagnosis, is a further example. In general science, risk assessment and prognosis correspond to *prediction*.

The second major sub-category of Opinion corresponds to opinions about the *desired* state of the patient system, and includes much of the creative thinking of the clinical investigator. Once a diagnosis, risk assessment or other evaluation about the patient's condition has been formulated, the clinical investigator determines what to do about it. Three key types of clinical thought at this point are *scenarios* (or what-if statements, *goals* and *plans*. A goal, such as weightloss and a target weight of 85kg, is a statement about *what* the desired state of the patient system should be, while a plan is a statement about *how* to get there.

In summary, the Opinion category is distinguished from the Observation category by representing inferences from evidence, rather than representing the evidence. Two investigators can form different interpretations of the same set of observations, but the observations themselves remain an objective picture of some aspect of the patient's situation, within the limits of the observational method itself. Similarly, two investigators can formulate different goals and plans based on the same observations, and even the same diagnosis.
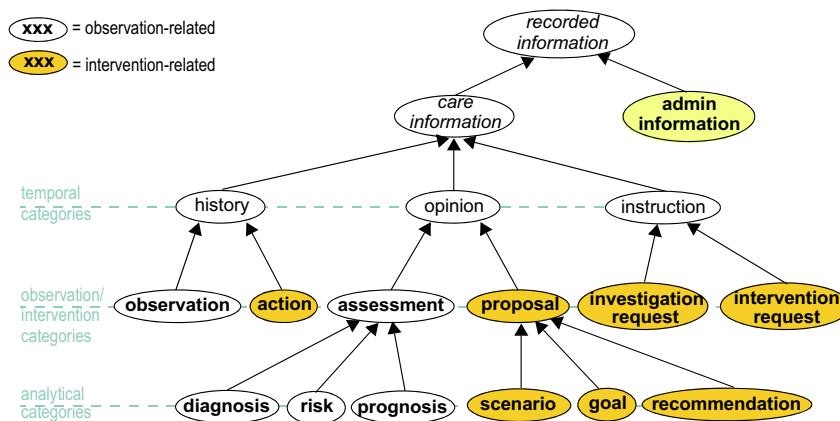
Figure 5 - The Clinical Investigator Record (CIR) Ontology

The clinical investigator record ontology

The ontology thus described is illustrated in Figure 5. Categories which correspond to the describing the patient system are shown using plain bubbles, while categories corresponding to interventions intended to change the system are shown using filled bubbles. We call this ontology the Clinical Investigator Record (CIR) ontology. The CIR ontology does not yet include any detailed subcategories beneath Admin information. Far less work has been done on this subject from an ontological point of view in health to date.

Conclusion

The Clinical Investigator Record Ontology provides the basis for the Entry classes in the *open*EHR reference model. The latter has proved to be semantically robust and a successful basis for defining clinical content models (known as 'archetypes') used in various kinds of health computing, as well as for EHR systems. The archetyping methodology has been used to create over 250 archetypes based on the *open*EHR Entry classes during the last few years, including in recent work (early 2007) in the UK NHS Connecting for Health programme.

References

[1] Heeks R, Mundy D, Salazar A. *Why Health Care Information Systems Succeed or Fail*. Information Systems for Public Sector Management - Working Paper Series Paper No. 9. 1999. Institute for Development Policy and Management.

[2] Scott J T, Rundall T G, Vogt T M, Hsu J. *Kaiser Permanente's experience of implementing an electronic medical record: a qualitative study*. BMJ 2005;331: 1313-1316.

[3] Sowa J. *Knowledge Representation: LogicalPhilosophical and Computational Foundations*. 2000 Brooks/Cole. (Top-level categories available at http://www.jfsowa.com/ontology/toplevel.htm).

[4] Rosse C, Kumar A, Mejino JLV, Cook DL, Detwiler LT, Smith B. *A Strategy for Improving and Integrating Biomedical Ontologies*. Proceedings of AMIA Symposium 2005, Washington DC, 639–643. (available at http://ontology.buffalo.edu/bio/OBR.pdf).

[5] SNOMED-CT. See http://www.snomed.org.

[6] Weed LL. Medical Records, Medical Education and Patient Care. *The Problem Oriented Medical Record as a Basic Tool*. Cleveland: Case Western Reserve University ress, 1968.

[7] Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: an analysis of clinical reasoning*. Cambridge, MA: Harvard University Press 1978.

[8] Rector A L, Nowlan W A, Kay S. *Foundations for an Electronic Medical Record*. The IMIA Yearbook of Medical Informatics 1992 (Eds. van Bemmel J, McRay A). Stuttgart Schattauer 1994.

[9] Bruun-Rasmussen M, Bernstein K, Vingtoft S, Nøhr C, Andersen SK. *Quality labelling and certification of Electronic Health Record Systems*. Studies in Health Technology and Informatics 2005; 116: p47-52.

[10] RICHE Consortium. RICHE ESPRIT Project. Final Report. Nov 30 1992.

[11] HL7 International. *Reference Information Model (RIM)*. See http://www.hl7.org.

[12] Van de Velde R, Degoulet P. *Clinical Information Systems: a Component-Based Approach*. 2003 Springer-Verlag New York.

[13] Beale T, Heard S. *The GEHR Object Model - Technical Requirements*. 2000. See http://www.openehr.org/gehr_australia/gehr_requirements.pdf.

[14] Beale T, Heard S. *Architectural Overview of openEHR*. 2006, The *open*EHR Foundation. See http://svn.openehr.org/specification/BRANCHES/Release-1.1-candidate/publishing/architecture/overview.pdf.

Address for correspondence

Thomas.Beale@OceanInformatics.biz.