*The New Navigators: from Professionals to Patients*
*R. Baud et al. (Eds.)*
*IOS Press, 2003*


# An ontological approach to evidence-based medicine and meta-analysis

**Domenico M Pisanelli[a], Davide Zaccagnini[b], Lucio Capurso[c], Maurizio Koch[c]**

[a] *National Research Council - Institute of Cognitive Science and Technology, Rome, Italy*
[b] *'Policlinico Umberto I' University Hospital, Rome, Italy*
[c] *'San Filippo Neri' Hospital, Rome, Italy*

**Abstract**

*The "evidence-based medicine" (EBM) paradigm is centered on the concept of "best evidence" and clinical studies based on this approach are more likely to be considered by physicians in their practice.*
*In this paper we describe an ontology representing the concepts involved in evidence-based medicine and meta-analysis and show how an ontological approach can be applied both for revisiting EBM conceptual foundations and for allowing a more effective knowledge-based information retrieval in literature.*

**Keywords:**
Ontology; Evidence-Based Medicine; Meta-Analysis

## 1. Introduction

The ever increasing amount of available information is definitely a strategic resource for today's physicians, but also a cumbersome matter to deal with. Knowledge-based information retrieval has been recognized as a prominent issue in medical informatics research field since many years ago. However, still a small portion of this information has a real impact in clinical environment.

Physicians are dealing with a rapidly changing, unstructured and qualitatively heterogeneous knowledge environment while trying to pick up the needed information for decision making [1]. It has been stated that a general physicians faces the task of examining 19 articles a day just in order to keep up with relevant medical journals [2].

A new paradigm has emerged in order to ensure the quality of knowledge reported in literature and, eventually, to reduce the "noise" of less relevant papers. Such paradigm is centered on the concept of "best evidence" and clinical studies based on this approach are more likely to be considered by physicians in their practice.

The evidence-based medicine (EBM) has been defined in 1997 by Sackett as "the conscientious, explicit and judicious use of current best evidence in making decision about the care of the individual patient" [3]. One of its principal scopes is thefore to provide health professionals with useful information in order to offer the best available medical procedures to patients and improve their clinical outcome.

Applying issues of scientific method theory to clinical research has been the theoretical revolution made by EBM doctrine in order to make medical knowledge more structured and reliable. This has supposed to be achieved by using best epidemiological and statistical methods in clinical studies (trials) design and execution. Information obtained from these experimentally controlled sources should be then transported and used by doctors while caring for patients.

Selecting and utilizing the evidence is thus a life long practice made by clinicians seeking important clinical information about diagnosis, treatment, prognosis and other clinical issues. Clinical assumptions based on "evidence" should modify physicians behavior restricting unjustified variability, clinical entropy, and guaranteeing quality standards between health care providers [4].

Still, debate is continuing on conceptual and practical aspects of the EBM movement.

Many authors underlined critical points in applying evidence-based medical knowledge in real world clinical contexts [5]. This is presumably caused by EBM studies' methodology which is certainly able to provide solid answers on selected topics of clinical interest, but forces scientists to remarkably simplify or ignore the natural variability of patient's presentation and behaviors.

A fertile debate recently arose about increasingly sophisticated statistical inferences as method for answer to clinical questions [6]. Quantitative meta-analysis for instance have been told to be not completely a reliable mean to medical knowledge building since different results have been obtained by these means about same clinical issues. Practical clinical guidelines developed by different professional bodies are sometimes conflicting.

On a general view both pragmatic and theoretical arguments are limiting EBM's expected diffusion in medical practice.

## 2. The advantages of an ontological approach

Physicians developed their language in order to achieve an efficient way to store and communicate general medical knowledge and patient-related information. This language was appropriate for the only support available for archiving, processing and transmitting knowledge: paper.

Paper-based terminology systems cannot satisfy any longer the new desiderata of healthcare information systems, such as the demand for re-use and sharing of patient data, their transmission and the need of semantic-based criteria for purposive statistical aggregation. The unambiguous communication of complex and detailed medical concepts is now a crucial feature of medical information systems.

How to cope with a clinical study in which some records report patients' temperature in Celsius degrees and some other ones in Fahrenheit? Apart from the undesired lack of standardization, it would be trivial to treat data mismatch in such a case.

But what happens when less evident data heterogeneity is found into aggregated records? There is the risk of not taking it into account, letting the reliability of knowledge extracted from clinical data decrease.

When an ontological approach is adopted in modeling the processes of data collection and aggregation, the tacit knowledge which is frequently source of potential misunderstandings among experts is made explicit.

Such an ontological approach is based on enumerating all concepts used in a domain and in providing their formal definitions according to suitable formalisms (usually logic-based).

Apart from its definition in a philosophical context - where it refers to the subject of existence - *ontology* in information science is "a partial specification of a conceptualization", as stated by Gruber [7], whereas Sowa proposed the following definition influenced by Leibniz [8]:

> "The subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called an ontology, is a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D."

Actually there is some disagreement on what is an ontology. Some admit informal descriptions and hierarchies, only aimed at organizing some uses of natural language; others require that an ontology be a *theory*, i.e. a formal vocabulary with axioms defined on such vocabulary, possibly with the help of some axiom schema, as in description logics.

In our perspective, an ontology is a formal theory which specifies the conceptualization (i.e. the intended meaning) of a lexical item as it is used in a certain domain. Since lexical items are often used with more than one meaning in the same domain (they are "polysemous"), such different conceptualizations have to be specified and different names must be given to concepts in order to distinguish their meanings.

For example, the lexical item "meta-analysis" – according to the context in which is used – may stand both for "analytical activity performed" and for "results obtained".

Ontologies therefore may help designing unambiguous information systems. If "no man is an island", "no system is an island" anymore: data and knowledge integration are no longer an optional, but a clear necessity. In fact, the overwhelming amount of information stored in various data repositories - including the web - emphasizes the relevance of knowledge integration methodologies and techniques to facilitate data sharing.

The need for such integration has been already perceived for several years, but telecommunications and networking are quickly and dramatically changing the scenario.

Even in medical information systems, ontologies are gaining an ever increasing role and several research activities are carried on in this field (see for example the work on UMLS Metathesaurus [9] and the study on genetic ontologies [10]). Times are ripening for designing an ontology aiming at representing the concepts involved in evidence-based medicine and meta-analysis.

## 3. The ontology of evidence-based medicine and meta-analysis

The role of ontologies is that of representing the conceptualizations behind any information and knowledge processing activity.

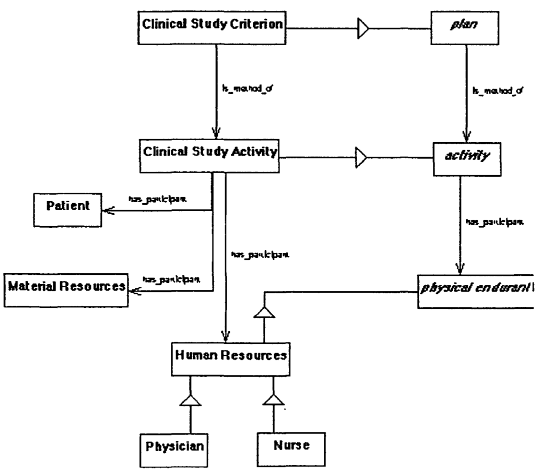


Figure 1
The main concepts in our ontology related to "Clinical Study Activity" (arrows stand for relationships, triangles represent IS_A relationships).

In the context of this research, we center our ontology design on the concept of "clinical study activity" (figure 1). Every clinical study is based on a certain criterion, whether or not made explicit. Therefore we define the concept "clinical study criterion" which "is_method_of" "clinical study activity". In the figures we represent concepts in boxes and relationships with arrows, triangles stand for IS_A relationship (concept-A IS_A concept-B means that every instance of A is instance of B, e.g. "feline" IS_A "mammal").

In developing this domain ontology, we re-use some concepts from domain-independent higher level ontologies. In particular from the plan and activities ontology in which it is stated that "plan" "is_method_of" "activity" and "activity" "has_participant" "physical endurant" (i.e. a physical object).

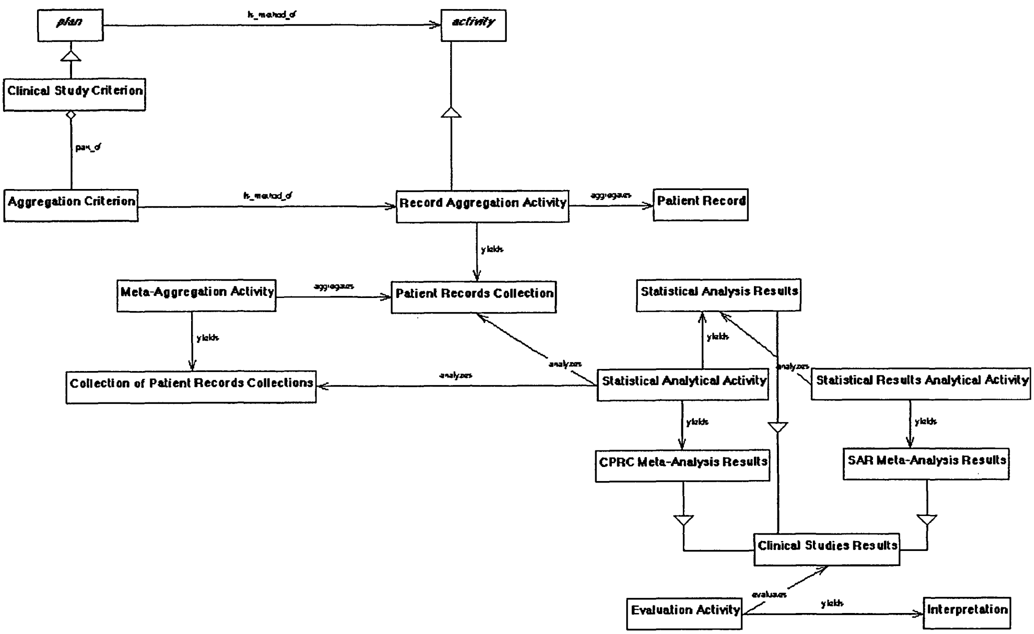


Figure 2
Collection of patient records and collection of collections: basic features of aggregation and meta-aggregation and two kinds of meta-analyses distinguished.

Such high-level concepts are taken from the DOLCE foundational ontology [11] which contains a description of the basic kinds of entities and relationships that are assumed to exist in some domain, such as process, object, time, part, location, etc. The role of DOLCE is not to state how things are, but to help categorizing an already formed conceptualization and to represent it according to some existing knowledge.It is profitable for domain experts to validate the main concepts involved in an ontology by means of graphical sketches, however an ontology is represented by a proper formalism (e.g. Ontolingua [12]) or a description-logic based language like Loom [13] that we employ.

In fact, graphics is not enough to represent the complexity of concepts and relationships among them. In figure 2 we sketch out the static relationships involved in the record data aggregation process. An aggregation criterion is method of the record aggregation activity which performs aggregation on patient records and yields a collection of patient records. This is represented in Loom as:

```
(defconcept Record-Aggregation-Activity
      :is-primitive (: and Activity
                              (:some aggregates Patient-Record)
                              (: some yields Patient-Record-Collection)))
```

The picture alone is not able to account for all the aspects of this ontology module. We report its main features in natural language, although a formalized version in Loom exists.

The collection of patient records (pertaining a single health-care structure) can be furtherly aggregated, together with other collections, yielding a "collection of collections" (those data employed in meta-analysis).

A statistical analysis (in the sense of "analytical activity") can be performed both on the simple records' collection and on the collection of collections. In the first case we obtain ordinary statistical analysis results, in the second one we get meta-analysis results that we name CPRC (collection of patient record collection) results in order to distinguish them from another class of meta-analysis results.

In fact CPRC results are obtained from an analysis on originally disaggregated data, whereas there might be the case in which several clinical studies are put together, but only statistical data for each study are put together and examined. In this case, researchers analyze results of statistical analyses and not the original data. We name the results of such activity "SAR (statistical analysis results) meta-analysis results".

Last but not least, it must be mentioned that every clinical study is originated by a clinical study hypothesis, and results are evaluated by researchers and interpreted, contributing, usually in the form of scientific journal papers, to the advancement of medical knowledge.

## 4. Applying the ontology

Apart from the trivial application of this ontology in the didactical field, where it could be used to introduce fundamentals of EBM and meta-analysis to sudents, the EBM ontology has at least two distinct application potentials.

From a more abstract point of view, the EBM definition itself could be revisited - also taking profit from the ontological formalization of its current paradigm - in order to broaden its scope. Medical science needs a wider perspective of what "evidence" is. There are entities in EBM current conceptual framework escaping clear definition, but still substantially concurring to EBM physiology and dynamics. A better understanding of this unexplored features will add concreteness and flexibility to EBM practice.

On the other hand, from a more practical point of view, we are designing an immediate application of this ontology. In fact, the concepts singled out and formalized by us are the groundwork for implementing an information retrieval system more advanced than an ordinary one based on keywords.

Let us take, as an example, the following text taken from the literature: "To assess the association between radon exposure and lung cancer, we conducted a population-based case-control study of Iowa women aged 40 to 84 who lived in their current home for at least 20 years." [14] If we map this sentence into our ontology, the first part (till comma) is a "clinical study hypotesis", the second part is a "clinical study criterion".

Let us examine a more complex example, taken from the summary of a cohort study [15]:

> *Background*
> Hyperthermia acts as a teratogen in some animals where it can induce resorption of the fetus and fetal death. [...]
> *Methods*
> We interviewed 24.040 women who were recruited in the first half of pregnancy to the Danish National Birth Cohort Study, and obtained information on the number of fever incidents during the first 16 weeks of pregnancy. [...]
> *Findings*
> 1145 pregnancies resulted in a miscarriage or stillbirth (4·8%). During the first 16 pregnancy weeks 18·5% of the women experienced at least one episode of fever. [...]

*Interpretation*

We found no evidence that fever in the first 16 weeks of pregnancy is associated with the risk of fetal death in clinically recognised pregnancies.

In this case "background" is mapped into "clinical study hypotesis"; "methods" must be furtherly analyzed, since it refers both to "clinical study criterion" and "clinical study activity" (to be furtherly decomposed into sub-criteria and sub-activities); "findings" is mapped into "SAR Meta-Analysis Results" (see § 3) and "interpretation" is "interpretation" in our ontology.

The implementation of such an ontology-based information retrieval system is still a work in progress. We envisage its application for a more effective data mining in medical literature. We plan to profit from natural-language understanding techniques in order to perform semi-automated knowledge-based tagging of scientific papers.

**References**

[1] Strauss S E, Sackett D L, "Using research finding in clinical practice", BMJ 1998;317:339-342.
[2] Davidoff F, Haynes R B, Sackett D L, Smith R., "Evidence-based medicine: what it is and what it isn't", BMJ 1995;310:1085-1086.
[3] Guyatt G H, Sackett D L, Sinclair J C, Hayward R, Cook D J , Cook R J., "Users' guides to the medical literature. IX.", JAMA 1995 Dec 13; 274(22):1800-1804.
[4] Haynes B, Haines A., "Barriers and bridges to evidence based clinical practice", BMJ 1998; 317: 273-276.
[5] Sackett D L, Haynes R B., "The architecture of diagnostic research", BMJ 2002; 324: 539-541.
[6] Malterud K, "Reflexivity and metapositions: strategies for appraisal of clinical evidence", J Ev Clin Pr. 2001; 8, 2: 121-126.
[7] Guarino N, "Formal Ontology and Information Systems", in Guarino N (ed.), Formal Ontology in Information Systems, Amsterdam, IOS-Press, 1998.
[8] Sowa J, communication to the ontology-std mailing list, 1997.
[9] Pisanelli DM, Gangemi A, Steve G, "A Medical Ontology Library that Integrates the UMLS Metathesaurus™", LNAI 1620, Berlin, Springer Verlag, pp. 239-248, 1999.
[10] Pisanelli DM, Battaglia M, Gangemi A, Steve G, "Ontological analysis for the unification of biology", JAMIA Symposium Supplement, 2002.
[11] Masolo, C., Gangemi, A., Guarino, N., Oltramari, A., and Schneider, L.: WonderWeb Deliv. D17: The WonderWeb Library of Foundational Ontologies, 2002.
[12] Gruber T. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition; 5:188-220, 1993.
[13] MacGregor RM, "A Description Classifier for the Predicate Calculus" Proc. AAAI 94.
[14] Field, RW, Lynch, CF, Steck, DJ, Smith, BJ, Brus, CP, Neuberger, JS, Woolson, RF, Fisher, EF Platz,CE, Robinson, RA, "Iowa Radon Lung Cancer Study Abbreviated Methodology", on line at: http://www.cheec.uiowa.edu/misc/radon.html.
[15] Nybo Andersen AM, Vastrup P, Wohlfahrt J, Andersen PK, Olsen J, Melbye M, "Fever in pregnancy and risk of fetal death: a cohort study", Lancet online 2002; http://image.thelancet.com/extras/01art12058web.pdf

**Address for correspondence**

Domenico M. Pisanelli, CNR - ISTC, Viale Marx 15, 00137 Rome, Italy
pisanelli@acm.org http://ontology.ip.rm.cnr.it