Available online at www.sciencedirect.com
ScienceDirect
Web Semantics: Science, Services and Agents on the World Wide Web 4 (2006) 229–236
JOURNAL OF Web Semantics
www.elsevier.com/locate/websem

Referent tracking for treatment optimisation in schizophrenic patients: A case study in applying philosophical ontology to diagnostic algorithms

Werner Ceusters [a,b,*], Barry Smith [c,d]

[a] New York State Center of Excellence in Bioinformatics and Life Sciences, 700 Ellicott Street, Buffalo, NY 14203, USA
[b] Department of Psychiatry, University at Buffalo, Buffalo, NY, USA
[c] Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany
[d] Department of Philosophy, University at Buffalo, Buffalo, NY, USA

Received 15 February 2006; accepted 1 May 2006

Abstract

The IPAP Schizophrenia Algorithm was originally designed in the form of a flow-chart to help physicians optimise the treatment of schizophrenic patients in the spirit of guideline-based medicine. We take this algorithm as our starting point in investigating how artifacts of this sort can benefit from the facilities of high-quality ontologies. The IPAP algorithm exists thus far only in a form suitable for use by human beings. We draw on the resources of Basic Formal Ontology (BFO) in order to show how such an algorithm can be enhanced in such a way that it can be used in Semantic Web and related applications. We found that BFO provides a framework that is able to capture in a rigorous way all the types of entities represented in the IPAP Schizophrenia Algorithm in way which yields a computational tool that can be used by software agents to perform monitoring and control of schizophrenic patients. We discuss the issues involved in building an application ontology for this purpose, issues which are important for any Semantic Web application in the life science and healthcare domains.

Keywords: IPAP; Schizophrenia Algorithm; Realist ontology; Basic Formal Ontology; Electronic health record; Referent tracking; Automated decision support

1. Introduction

The International Psychopharmacology Algorithm Project (IPAP) is an internationally supported initiative set up in 1985 by a team of psychiatrists, psychopharmacologists and algorithm designers in an effort to improve choice of medication in psychiatry [1]. In 1995 [2], the *IPAP Schizophrenia Algorithm* (IPAP-SA) was published by IPAP as a guideline consisting of four schizophrenia treatment algorithms developed, respectively, for the first schizophrenic episode, long-term medication maintenance, schizophrenia complicated by comorbid psychiatric disorders, and schizophrenia complicated by neuroleptic malignant syndrome. In January 2005, the latest version of this IPAP guideline (v.20041223) was released and made available on the web [3]. The algorithm is presented in the form of a flow-chart, which has one entry condition – namely an established diagnosis of schizophrenia or schizoaffective disorder – and two exit conditions, one suggesting a modification to the patient's current treatment program, the other suggesting unaltered continuation of this program. The algorithm consists of instructions like: 'if the patient exhibits severe agitation, consider a 4–6-week trial of an atypical or, if not available, a trial of a typical antipsychotic'. The nodes in the chart represent either *questions* related to the condition of the patient or *instructions* on how to modify the existing treatment regimen. In the electronic, hyperlinked version of the flow-chart, the first set of nodes point the user in the direction of more detailed specifications of symptoms to look for; nodes in the second set point to drugs which might be prescribed and to scientific papers motivating treatment suggestions. The whole is also available in a pdf format that can be printed out and consulted when the clinician is not on-line.

The on-line version provides some obvious advantages over a traditional journal or textbook publication. It can be accessed immediately through any suitable browser, and new versions become accessible as soon as they are released. On the other

* Corresponding author at: New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Washington Street, Buffalo, NY 14203, USA. Tel.: +1 716 563 6663.
E-mail address: ceusters@buffalo.edu (W. Ceusters).

1570-8268/$ – see front matter
doi:10.1016/j.websem.2006.05.002

side, however, given that the algorithm is currently implemented as a simple flow-chart in which the included hyperlinks serve only human browsing, it still fails to exploit the real power of the computer, which is to perform reasoning automatically. An implementation could for instance draw on information already available in the patient's electronic health record (EHR) in such a way as to process relevant features of the patient's current condition in light of those criteria which play a role in the corresponding step of the algorithm. A novel approach for building such implementations is that of *referent tracking*, which relies on an approach to the representation of entities in the healthcare domain that is based on philosophical realism [4]. The cornerstone of this approach is the distinction between particulars (such as John, or the fracture in his right tibia), and the universals (*human being*, *fracture*, *tibia*) which these particulars instantiate. The purpose of the research reported here was to carry out the first step of enhancing the present version of the IPAP algorithm in such a way that it can be used in automatic decision support. To this end it is necessary to identify the minimal set of universals and particulars (tokens, instances) which must be represented in a referent tracking system in order to allow software agents to carry out real-time monitoring and control activities to optimise the treatment of schizophrenic patients in accordance with IPAP guidelines. The analysis performed is intended to be a case study with the goal of demonstrating how this approach could be used for upgrading similar static and inert flow-chart algorithms in such a way that they would constitute dynamic application ontologies.

2. Background

The system we advance as a novel enhancement of the IPAP-SA falls under the paradigm of 'expert system for real-time monitoring and control' that is described in [5]. However, it is marked also by some special features of its own.

First, the IPAP-SA is designed to support monitoring and control primarily on the basis of phenomena observed on the side of the patient which are *qualitative*, in the sense that they cannot be associated with quantitative magnitudes (in some cases because quantitative values cannot be measured by means of any currently available device). Thus, the algorithm embodies a qualitative ontology of the domain of schizophrenia and its treatment.

The second distinguishing feature concerns the issue of *when* the system we envisage should be invoked. Traditionally, questions are raised in the course of using an expert system only when [1] the system has already been activated by the clinician in the course of a clinical encounter to assist him in his decision making and [2] the inference engine encounters a node where 'if-then'-conditions need to be evaluated in order to determine the next step to be taken. Under the paradigm advocated here, in contrast, the IPAP-SA would be embedded in a monitoring system that would keep track continuously and in real time of changes in those electronic health records to which the monitoring system has access. The execution of the algorithm would then be triggered automatically by relevant changes in the EHRs associated with each patient. Note that these need not be changes in the patient's own EHR. In the case of *significant loss*, for example, a term used in psychiatry to indicate that the death of a loved one may result in psychiatric disturbances, the event in question will be one that is registered initially in the EHR of some other person. In the scenario we envisage, the registration of such an event would trigger a software agent to initiate a process designed to establish whether the event does indeed constitute a significant loss for some other person under scrutiny. Clinicians or patients may then be called upon by the system to provide answers to additional questions generated by the software, answers which cannot be found automatically in the relevant EHR data. An obvious benefit of this approach, which is known as the *mixed initiative interaction paradigm*, is the more accurate management of risk. Since, for a schizophrenic patient, a significant loss is qualified as a potentially dangerous situation that might lead to suicidal ideation, one should try to anticipate such ideation by contacting the patient with a view to further monitoring and possibly modifying his treatment as soon as possible after the significant loss, given that the time of the next scheduled visit may already be too late.

The third distinguishing feature concerns the origin of the data on which the algorithm will operate. For most expert systems for monitoring and control, the data derive from some single patient or from the patient's immediate environment (e.g. the temperature or air humidity in an IC unit). The IPAP Schizophrenia Algorithm, in contrast, rests on information that may need to be looked for elsewhere, including in other information systems containing EHRs, or in the wider environment of the patient. This requires the use of a *referent tracking system*.

2.1. Referent tracking

No EHR can ever be complete. Yet it is often the case that a plurality of EHRs are maintained by different institutions for the very same patient, often with considerable overlap and complementarity of data, but sometimes also with contradictory information. In addition, as argued above, data in one patient's EHR may be relevant to the care of other patients.

Unfortunately, however, the information contained in EHRs is typically stored in such a way that it is difficult to tell whether it is consistent, or whether given data refer to the same or to distinct entities in reality. To resolve these problems we have proposed that EHRs should be created in such a way that both software and human agents can know precisely what included statements are about. This problem is solved for individual patients via the mechanism of unique alphanumeric patient identifiers. On our proposal this same mechanism would be extended to all the particulars to which reference is made in the clinical record [4]. Systems containing data pertaining to the same particulars could then be connected automatically, so that relevant changes in one system come immediately to the attention of the other systems. The realization of this goal will require a more dynamic type of connectivity than that which has been realized hitherto in the context of patient data warehouses or standard EHR messaging—but connectivity of this sort will in any case increasingly be needed by healthcare institutions, and it will be needed also to take full advantage of the ongoing trend for patients to maintain a personal health record on the Internet, thereby keeping track of their own medical history [6].

In [4], we have sought to provide arguments to the effect that expert systems created for real-time monitoring and control should exploit the facilities of a common referent tracking system that is designed to guarantee that each of the real-world entities found to be of salience in the course of the medical care of each given patient would be individually represented by its own explicit (alphanumeric) identifier in that patient's EHR. Current EHRs consist almost exclusively of statements to the effect that a given patient has been diagnosed with a disorder of such and such a type or has received a treatment of such and such a type from this or that physician, where only the patient and the physician involved receive unique identifiers. All other information in the EHR is conveyed through combinations of time-stamping information with general codes from clinical terminologies (as in "John Smith's schizophrenia, first diagnosed at 4 p.m. on 2006/03/29").

In the type of EHR we have in mind, in contrast, this particular schizophrenia – which is to say, this particular *case* or *instance* of this disease in this particular patient – would also receive its own unique alphanumeric identifier, hereafter called as "Instance Unique Identifier" (IUI). In this way it becomes possible to include within an EHR statements which are directly about what it is on the side of the patient to which we refer when talking about each particular instance. The assignment of IUIs should be systematic—thus IUIs should be applied (wherever possible by software) to all instances salient to the healthcare of a given patient at the point where corresponding general codes are entered into the clinical record of that patient, which means not only to instances of disorders, but also to the patient's body parts (his brain, a specific region of his brain, a specific tumour), particular episodes in his clinical history, particular symptoms he presents, particular family members, and so on. The feeling of distress that he experiences today will receive a new IUI, distinct from the IUI received by the similar feeling he experienced 2 years ago, even though both are associated with the same ICD10 or DSM-IV code.

The resultant *referent tracking database* (RTDB) may form part of the legacy EHR system of the specific institution in which the patient is managed. Better, however, would be to have the computer systems of pluralities of institutions linked together systematically in such a way that updates in referent tracking data deposited in one system would immediately trigger relevant actions in the others.

IUIs can be used in statements asserting that given particulars are instances of given universals, for example, that particular #12345 is an instance of the universal *schizophrenia*. Or they can be used in statements asserting that given particulars stand in given relations to each other, for example, that particular #12345 is *contained_in* particular #12300. Ideally, relations will be used which are defined within a formal framework of the type outlined in [7]. The encoded statements could then serve as basis for logical reasoning, for example, in Semantic Web applications. It is for this reason that the formal specification of the RTDB is expressed in a tuple-oriented language, so that statements therein can be easily translated into RDF or OWL [8]. IUIs can be used to track particulars whose nature changes over given periods of time, and also for reasoning about particulars whose nature is unknown. Thus, we can assign a IUI when a disorder is postulated as being responsible for a given family of symptoms even before we know enough to classify the disorder in a determinate way.

3. Hypothesis

Our basic idea is that a patient's EHR would be associated with pertinent algorithms in such a way that changes in the RTDB would trigger the automatic execution of those steps in patient care which the algorithms recommend (steps which sometimes need to be taken urgently, including the issuing of relevant warnings). In order to assess whether it would be possible to use the IPAP-SA in a setting of this sort, it was necessary to subject it to a terminological and ontological analysis with the goal of identifying any obstacles it might present to automation along the lines described. The first such potential obstacle would be the failure to trace back all the data elements mentioned in the algorithm to corresponding particulars or universals of the sorts which are recorded in a referent tracking system. Our hypothesis was that this effort could be conducted successfully.

4. Methods

For our analysis, we used the version (v.20041223) of the IPAP-SA available on the IPAP website on June 23, 2005 [1]. We studied in detail the explanations provided for each of the nodes in the flow-chart on the basis of the information sources referenced by the algorithm itself. We made an inventory of the data elements which would be needed in order to assess whether a given patient's condition fulfils the criteria advanced in the decisional nodes of the flow-chart, identifying thereby also some of the resources (including other relevant guidelines) that would be required in making accurate assessments and comparisons of the corresponding entities in reality. We then classified these data elements according to whether they pertain to the realm of particulars (e.g. this patient's headache here and now) or of the universals which these particulars instantiate (*patient*, *headache*, *death*, *significant loss*, etc.). We evaluated the latter according to the theory of Basic Formal Ontology (BFO) presented in [9], which enforces distinctions such as that between enduring entities (such as the patient's brain) – also called *continuants* – and associated processes (e.g. the tremor of his left lower arm)—also called *occurrents*. It enforces also the distinction between independent entities (the patient, his arm, his brain) and dependent entities (his headache, age, or temperature), the former serving as bearers or carriers for the latter.

If a given patient exists, then certain other particulars necessarily exist also (or have existed at some earlier time). As an example: if a given patient exists, then his age necessarily exists—there are no persons without an age. This age is a particular quality to which (if such an assignment has not already been made) it is possible to *assign* a IUI. In the case of attributes like age, which vary continuously through time, we need to distinguish the attribute itself from those *values* which the attribute takes on at successive moments in time (which will not be assigned IUIs of their own). In the case of variable attributes such as temperature, corresponding numerical data will be stored with time-stamp information in the RTDB.

In contrast to what is the case for age or temperature, disorders are not such that, given the existence of the patient, some particular instance of disorder, too, must necessarily exist. This means that a term like "catatonia" as used in the algorithm refers not to some determinate particular but rather to a universal which may or may not be instantiated in this particular patient. Thus, when a referent tracking system is being used in support of patient evaluation via the IPAP-SA, it must be determined through clinical assessment by physician or nurse whether there exists, for each such universal, a corresponding instantiation on the side of the patient.

To maximize the benefits to be derived from using a referent tracking system in the way described in the foregoing, we must ensure that the various entities under review are treated in appropriate ways by the system in light of the ontological categories to which they belong. To this end we specified for each particular what BFO category it is an instance of and for each universal by what top-level BFO category it is subsumed.

5. Results and discussion

While all nodes in the IPAP-SA could be expressed by means of one or other BFO category (or some combination thereof), this was not in every case achieved without difficulty. Indeed, identifying the ontological status of many of the entities referred to in the algorithm required a detailed prior analysis. We believe that this analysis may provide useful lessons for those embarking on similar endeavors in the future and thus we provide a number of examples in what follows, focusing here on those nodes which are of relevance to the life science and healthcare domains taken quite generally.

5.1. Persons and patients

The most important particular, from the perspective of the IPAP-SA, is the person to be monitored. Assigning IUIs to persons is of course common practice in many administrative contexts, though their 'uniqueness' is often restricted to the environment in which a given information system (whether electronic or manual) operates. The library from which you borrow books, the social security program in which you are registered, the company for which you work, the health facility in which you once were treated—all will likely have their own more or less idiosyncratic systems of personal IDs. In the domain of healthcare, however, the fact that the very same person might be supplied with more than one unique identifier hampers the aggregation of pertinent data in a way which increases costs and the likelihood of error [10], so that many countries have initiated programs to implement systems of unique patient IDs that are valid nationwide, though in some countries (as in the case of the US), such a measure is still seen as controversial because of privacy concerns [11].

From an ontological perspective, it is important to note that we envisage assigning an ID to the patient as *person* and not as *patient*. A person is an entity that starts to exist some time at or around conception, and ceases to exist when it dies. This entity is an independent continuant, and thus has to be distinguished from other entities with which it is closely associated, such as the person's *life* (a dependent occurrent). The term "patient", on the other hand, refers obliquely to entities which are to be classified as dependent continuants, namely to the *roles* which persons start to play, for example, when they seek certain sorts of advice in relation to their health. This same term is however also used in a second sense, as signifying *person who plays the patient role*. This tells us that, while natural language can be used to establish the ontological nature of the entities referred to by its means, it should not be trusted to give reliable indications in this respect in every case.

5.2. Diseases

The second most important particular, from the perspective of the IPAP-SA, is the disease that is being tracked. Diseases are dependent continuants. A disease is dependent because it cannot exist without some other entity (an organism) in which it exists: without human bodies there would be no human diseases. The thesis that a disease is a continuant is more difficult to grasp. Some consider a disease to be a process, and hence an occurrent [12]. The BFO view is that a disease is a *disposition* or *power* analogous to a function (e.g. of a bodily organ) but classified rather as a *dysfunction*, of a sort which affects or will affect the well-being of the organism. The disease comes into existence at a certain time and thereafter gives rise to a complex series of processes (such as manifestations of symptoms), which together form the *course* of the disease and can be recorded in the disease *history*. It may give rise also to associated continuants (a tumour, a redness of the skin, . . .), which come into existence as a result of the processes in question and depend for their continued existence on their further continuation [13]. The relevant disposition or power can be instigated by a bodily aggressor such as a virus, by biomolecular reactions that go wrong, or by processes called upon by the body to defend itself against something that is going wrong in the external environment in which the person lives.

We use the term *disease* in a sense both wider and narrower than that which is customary in the medical literature. It is wider because it comprehends relevant continuants (which may begin to exist as disturbances on the molecular level of granularity) even before they are detectable by any means available to human beings. It is narrower because it refers exclusively to dispositional entities (powers) and thus not, for example, to neoplasms, malformities and other pathological independent continuants which are commonly classified as diseases in many terminologies. We thus reserve the term "disease" for the disposition itself, which can evolve over time and undergo changes while preserving its identity. We use the term "disorder" to refer to anything through which the disease is manifested. The latter does not represent a single universal, since manifestations can be of very different sorts, including not only continuants such as *tumours* and *fractures*, but also processes such as *tremors* and *coughing*.

None of the latter examples are properly to be considered as diseases.

5.3. Schizophrenia

For the execution of the IPAP Schizophrenia Algorithm, it is mandatory that the patient's disease be an instance of one or other of the universals *schizophrenia* or *schizoaffective disorder*. We assumed thus far that such a determination has already been made. Now however we need to point out that the way the entry conditions are phrased in the algorithm as currently formulated poses certain problems. When we examine more closely the associated documentation on this topic, two important questions arise.

The first is logical in nature: does the patient's established diagnosis need to satisfy the diagnostic criteria of *both* DSM-IV and ICD-10, or is it sufficient that either one or the other be satisfied? This question is important, since there is only a partial concordance between the two, concrete figures for this concordance ranging from 60% (depending on the subtype of schizophrenia) to 83% [14]. Thus, it is possible that a patient's disease has to be classified as schizophrenia according to one system, but that it is not allowed to be so classified by the other.

The second question is ontological in nature: to what extent do the terms ("schizophrenia" or "schizoaffective disorder") used by ICD-10 and DSM-IV represent one, or two, or no universals at all on the side of biomedical reality? Here, too, the referent tracking idea might bring certain advantages. We first make what seems to us to be a reasonable assumption to the effect that, if a given body of patient records systematically includes diagnoses of schizophrenia and/or of schizoaffective disorder, then there is something to which these terms refer on the side of the corresponding patients. Each such something can be given a IUI, say #I-9001 for the schizophrenia diagnosed in John, and included in a referent tracking database that is used in carrying out a variety of different types of diagnostic tests. By analysing the results of such tests, we may in the long run be led to the conclusion that #I-9001 is in fact a compound of two or more disease particulars (or, in the worst case, that it is an empty ID designating no disease at all) [15]. In this way, in the course of time, experience might prove that "schizophrenia" itself is a term that has no referent, for example, because what had been thought to be a single disease is in fact a compound of several diseases hitherto not cleanly separated—in ways which might then lead to modifications to the IPAP algorithm itself.

5.4. The notion of 'risk'

It is a matter of debate whether or not we should accept the existence of a particular called "the patient's risk or possibility of committing suicide". Krimsky gives risk a metaphysical status as a unitary property or quality in the physical world [16]. We think this assumption calls for further research. One possibility is that suicide risk is something measurable that depends on the person, similar to his height or body temperature. Another would be that suicide risk is best analyzed as a property not of one single person but rather of a collection of persons. After all, a collection of persons is a particular in its own right. Suicide risk might also be analogous, for example, to the risk of suffering from leptospirosis. The latter depends almost entirely on the environment [17], and can be changed even though nothing changes in the persons themselves. This latter view places suicide risk in the realm of what mathematicians call *latent variables*, which are defined as unobservable attributes that causally influence observable behaviour and which are studied by statistically relating covariation between observed variables to the latent variables through the analysis of interindividual differences [18].

5.5. Gender

A patient's sex is one of the parameters determining suicide risk and thus relevant in the context of the IPAP-SA. There are several issues which arise when we attempt to work out what sort of dependent entity (quality, attribute) is at issue here, not least because there are different ways in which the term "sex" is used in relevant clinical contexts, where the term can refer alternately to phenotypic sex, genotypic sex, gonadal sex, or administrative sex. Clearly, any particular person has at any time only one determinable phenotypic sex. The same can be said for gonadal sex. By *determinable*, we mean that, with the necessary technical equipment and skills, it would be possible to determine the exact sex of any given person. It can be what we traditionally understand as *male* or *female*, or even *mixed*. Quite often, one can find in patient databases the value "unknown". Clearly, however, this is not to be understood in the same way that we understand "male" or "female": it is very unlikely that for a person whose sex has been entered as "unknown" one has taken all measures to determine the relevant value and then came to the conclusion that it is of a type never seen before. Rather, this entry reflects the fact that one did not or could not determine the given person's sex, or that, for whatever reason, the relevant checkbox has not been filled in properly in the database. To include *unknown* as a different type of sex, alongside *male* and *female*, reflects a confusion between epistemology and ontology of a sort which has been shown to be pervasive in the development of biomedical terminologies [19].

A peculiar notion is that of *administrative sex*, whose application, depending on the community in which it is defined, is based not only on scientific grounds but also on political, ethical, and even religious considerations. An example are the different treatments of the right of gender self-identification, which can make it possible that the same person has a different administrative sex in Australia and in the US [20]. Although Rector et al. claim that 'administrative sex is definitely a thing' [21], we do not know what kind of thing it would be, though it is clear to us that it enjoys a totally different ontological relationship with the person who serves as its bearer than do biological types of sex.

Finally, there exists also the notion of *genotypic sex*, for which the situation may be more complex, for example, because of mosaicisms in which not all the cells of the body of a person exhibit the same chromosomal pattern [22]. From an ontological perspective, there is no problem in describing such a situation by using an appropriate set of foundational relationships to describe how chromosomes relate to cells, cells to a person's body, and so forth [23]. However, what names we choose to assign to conditions with different distributions of mosaicism is a matter of convention on the part of the scientific community.

5.6. Race

Also subject to political and ethical debate is the notion of race, although being a white Caucasian has been shown to increase a male's risk for suicide [24], if not directly, than at least by means of other observed differences within racial groups such as differences in the degree to which children are emotionally neglected [25]. Although today nearly all geneticists reject the idea that biological differences are tracked by racial and ethnic distinctions, the sequencing of the human genome has identified certain genetic variants associated with different frequencies of disease susceptibility, environmental response, and drug metabolism in different ethnic and racial populations as traditionally defined [26]. This does not constitute an argument for the acceptance of race as a natural kind, or, for example, to accept *Caucasian human* as a subtype of *human* in the same sense that *human* is a subtype of *mammal*. But it does motivate the recognition of ethnic or racial subpopulations, though of course marked by difficulties in identifying criteria that would allow any particular human to be correctly assigned to one or other of these populations.

5.7. Behaviour and functionings

The IPAP-SA uses at several points the terms "behaviour" and "functioning", referring to entities classified by BFO as aggregates of processes. A person's behaviour and (psychosocial) functioning may of course exist independently of the knowledge any observer might (or might not) have about the processes in question. But problems are raised nonetheless by certain aspects of our treatment of behaviour and functioning. A person can behave between time t_1 and t_2 in exactly the same way as between t_3 and t_4, yet his behaviour might be appropriate on the first occasion but not on the second. We take it that the term "behaviour" is used when one seeks to describe what persons *do* in a way that is neutral with respect to what persons are *supposed* to do (i.e. out of any context in relation to which specific behaviour might be qualified with respect to norms, for example, as *appropriate* or *inappropriate*, or as *aggressive* or *manic*). The use of the term "functioning", in contrast, presupposes that behaviour is evaluated in light of some *function* to be exercised. Thus, to talk of a person's *psychosocial functioning* presupposes that his behaviour is being evaluated in light of the *functions* which need to be realized if the person in question is to preserve his integrity as member of a community. It is thus astonishing that so many ontologies in the domain of healthcare and the life sciences do not make this important difference between a function and the execution thereof [27,28].

IPAP's use of the term "behaviour" to designate a ('neutral') aggregate of processes leads to a number of ontologically important questions. Is it, for instance, possible that a particular person has several stretches of behaviour going on in parallel during some given period of time? How, more generally, are we to assign IUIs to particular stretches of behaviour, for example, if behaviour of some given type ceases for a time and then reappears? As mentioned earlier, precise answers to such questions would not be important if the IPAP-SA were to be used in isolation, i.e. as activated by one clinician sitting in front of a single patient. But they are vital if we want to allow adequate automatic reasoning on the part of multi-agent monitoring systems in which information about multiple particulars is collected for subsequent use for a variety of purposes.

6. Conclusion

We presented a first analysis of the IPAP-SA from an ontological perspective, primarily in order to find the relevant particulars and universals that must be represented in a multi-agent computer environment in order to allow for a new type of automatic monitoring of schizophrenic patients that will enable optimisation of their pharmacotherapeutic management. At the same time, we wanted to provide some insight into the important differences that exist between the various entities that go to make the *what it is on the side of the patient* in virtue of which we are able to ascribe properties such as *being a person*, *being male*, *being Caucasian*, *being at risk for suicide*, and so forth. These expressions encode relationships of different ontological types, whose proper understanding is crucial to the building of ontologies that can be re-used in clinical algorithms and related tools. Rather than discussing each particular and universal thus identified, we focused on those cases that are central to the idea of referent tracking, or that are exemplary for the kind of confusion that is often exhibited in biomedical ontologies.

Our analysis in terms of particulars and universals is however only a first, though still important, step. A deeper ontological analysis would involve the axiomatisation, as a basis for implementation of a system for automatic reasoning, of the foundational relations that tie the relevant entities together.

Under currently prevailing paradigms, developing software programs that exhibit intelligent behaviour comes down to building a 'conceptual model' (a simplified digital simulacrum of some domain) that selects from the real-world relevant generic features organised in a way suitable for the fulfilment of the program's task. It is in this light that the term 'ontology' is used in the field of knowledge representation to refer to 'the specification of a conceptualisation' [29] or to 'that part of a domain model that excludes the instances, yet describes what they can be' [30]. To build an ontology along these lines, developers use a representation language such as one or other flavor of OWL that allows them to verify automatically whether their model is internally consistent. Often, adepts of this approach consider it to be irrelevant whether or not their model represents any reality on the side of the patient, just so long as the program behaves in the way it is supposed to behave. The problem is that models built in this way cannot be used for purposes other than those for which they were originally designed. Building such models is a labour-intensive process, and indeed still the main bottleneck in building knowledge-based systems [31]. The huge amounts of energy expended on such models in biomedical informatics thus far constitute, we believe, a mammoth waste of resources.

The realism-based approach to domain ontologies outlined in this paper, in contrast, promises more stable and more well-founded ontologies and thus offers better perspectives for reusability of the data analysed in their terms. It does this by providing an alternative to the strategy for ensuring reusability of ontologies adopted on the currently dominant paradigm, where it is assumed that to make ontologies reusable they must be 'mapped' one with another. This can occur either off-line, by having domain experts spend time finding the common elements in the ontologies to be combined [32], or dynamically, e.g. by having two agents, each equipped with his own model, interact [33] in a process of ontology negotiation. However, we know of no truly successful examples of either methodology having been applied in practice. Indeed the belief (almost never questioned) that they can be so applied rests on a prior assumption, which we believe to be false, namely that there is an underlying common understanding, a basic agreement, for example, among all relevant domain experts, concerning what the various terms in an ontology '*mean*' [34]. For on what should such agreements on meanings be based? What could serve, here, as benchmark? Perhaps *concepts* can serve this purpose—were it not for the fact that this very term is itself used to mean such a wide variety of different things (including: 'units of knowledge', 'units of thought', 'meanings of terms', and so forth), quite often in such a way that it is very hard to understand which meaning is precisely intended [35].

On the instance-based approach sketched in the above, in contrast, it becomes possible for ontologies to be reused by other software programs, not least in order to support the kind of mixed initiative interaction paradigm for multi-agent systems which we referred to above, in virtue of the fact that all the agents involved are able to keep track of the very same referents in an unproblematic way by appealing to the RTDB. The approach thereby provides the best possible measure for whether two (or more) ontologies are indeed compatible; for it allows the world itself to serve as benchmark as concerns whether two universals in an ontology are or are not identical, and it provides a clear strategy for ensuring that ontologies are built which are reusable. Not the least advantage of the referent tracking approach, therefore, is that it rests not on agreements concerning the meanings of terms or concerning associated 'concepts', but rather on the universals and particulars that populate the world [36].

Acknowledgments

The present paper was written under the auspices of the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the Network of Excellence in Semantic Interoperability and Data Mining in Biomedicine of the European Union, and the project "Forms of Life" sponsored by the Volkswagen Foundation.

References

[1] The IPAP Schizophrenia Algorithm, 2004, http://www.ipap.org/history.php (Last visited: August 5, 2005).

[2] K.J. Bender, Algorithm project provides guides to current knowledge, Psychiatric Times, vol. XIII, issue 6, June 1996, Electron. J., http://www.psychiatrictimes.com/p960625.html (Last visited: August 5, 2005).

[3] IPAP Schizophrenia Algorithm Project, http://www.ipap.org/schiz/index.php (Last visited: March 23, 2005).

[4] W. Ceusters, B. Smith, Referent tracking in electronic healthcare records, in: R. Engelbrecht, et al. (Eds.), Medical Informatics Europe, IOS Press, Amsterdam, 2005, pp. 71–76.

[5] A. Crespo (Ed.), Artificial Intelligence in Real Time Control, Elsevier, 1994.

[6] C. Denton, Will patients use electronic personal health records? Responses from a real-life experience, J. Healthcare Inform. Manage. 15 (3) (2001) 251–259.

[7] B. Smith, W. Ceusters, An ontology-based methodology for the migration of biomedical terminologies to electronic health records, AMIA, submitted for publication.

[8] W. Ceusters, B. Smith, Strategies for referent tracking in electronic health records, J. Biomed. Inform., in press.

[9] Basic Formal Ontology, http://www.ontology.buffalo.edu/bfo/BFO.pdf (Last visited: March 24, 2005).

[10] L. Fernandes, C. Lenson, J. Hewitt, J. Weber, J.A. Yamamoto, Medical Record Number Errors: A Cost of Doing Business?, http://www.madison-info.com/MedicalRecordNumberErrors.htm (Last visited: March 31, 2005).

[11] A. Richardson, McKesson Says Connectivity Key to National Health Info Network, Letter to Dr. Brailer, National Coordinator for Health Information Technology, U.S. Department of Health and Human Services, January 18, 2005, http://www.mckesson.com/feature_012105.html (Last visited: March 31, 2005).

[12] M.N. Liebman, Biomedical informatics: the future for drug development, Drug Discov. Today 7 (October (20 Suppl.)) (2002) S197–S203.

[13] I. Johansson, Pattern as an ontological category, in: N. Guarino (Ed.), Formal Ontology in Information Systems, IOS Press, 1998, pp. 86–94.

[14] A. Bertelsen, Schizophrenia and related disorders: experience with current diagnostic systems, Psychopathology 35 (March–June (2–3)) (2002) 89–93.

[15] B. Smith, From concepts to clinical reality: an essay on the benchmarking of biomedical terminologies, J. Biomed. Inform., in press.

[16] B. Legesse, The need for an integrated theoretical framework for understanding smallholder farmers' risk perceptions and risk responses: a review essay and some experiences from Ethiopia, Citing: S. Krimsky, The role of theory in risk studies, in: S. Krimsky, D. Golding (Eds.), Social Theories of Risk, Praeger, Westport, 1992.

[17] A. Nardone, I. Capek, G. Baranton, C. Campese, D. Postic, V. Vaillant, M. Lienard, J.C. Desenclos, Risk factors for leptospirosis in metropolitan France: results of a national case–control study, 1999–2000, Clin. Infect. Dis. 39 (September (5)) (2004) 751–753.

[18] R. Silva, R. Scheines, Generalized Measurement Models, Research Report CMU-CALD-04-101, May 16, 2005, http://www-2.cs.cmu.edu/~rbas/cald04-101.pdf (Last visited: June 12, 2005).

[19] O. Bodenreider, B. Smith, A. Burgun, The ontology-epistemology divide: a case study in medical terminology, in: A.C. Varzi, L. Vieu (Eds.), Proceedings of the Third International Conference on Formal Ontology in Information Systems (FOIS 2004), IOS Press, 2004, pp. 185–195, ACM Press, FOIS 2004, Torino, November 4–6, 2004.

[20] S.K. Milton, Top-level ontology: the problem with naturalism, in: Formal Ontology in Information Systems, IOS Press, Amsterdam, 2004, pp. 85–94.

[21] A. Rector, et al., Foundations of the Semantic Web: Ontology Engineering. Building Ontologies 1, http://www.cs.man.ac.uk/~rector/Modules/CS646-2004/Ontology-building-2004-Lect-2.ppt (Last visited: April 1, 2005).

[22] A.E. Covone, G. Voglino, R. Ravazzolo, Sex chromosome rearrangements leading to partial aneuploidies and mosaicisms: use of QF-PCR for detection and quantification of the involved cell lines, Int. J. Mol. Med. 14 (October (4)) (2004) 743–746.

[23] B. Smith, W. Ceusters, B. Klagges, J. Kohler, A. Kumar, J. Lomax, C.J. Mungall, F. Neuhaus, A. Rector, C. Rosse, Relations in biomedical ontologies, Genome Biol. 6 (5) (2005) R46 (Epub April 28, 2005).

[24] T.K. Pinhey, S.R. Millman, Asian/Pacific islander adolescent sexual orientation and suicide risk in Guam, Am. J. Public Health 94 (2004) 1204–1206.

[25] A. Roy, African American and Caucasian attempters compared for suicide risk factors: a preliminary study, Suicide Life Threat Behav. 33 (Winter (4)) (2003) 443–447.

[26] M.W. Foster, R.R. Sharp, Race, ethnicity, and genomics: social classifications as proxies of biological heterogeneity, Genome Res. 12 (June (6)) (2002) 844–850.

[27] B. Smith, J. Williams, S. Schulze-Kremer, The ontology of the gene ontology, AMIA Annu. Symp. Proc. (2003) 609–613.

[28] A. Kumar, B. Smith, The ontology of processes and functions: a study of the international classification of functioning, disability and health, AIME 2005 Workshop Biomedical Ontology Engineering, Aberdeen, July 24, 2005, in press.

[29] T.R. Gruber, A translation approach to portable ontologies, Knowl. Acquisition 5 (2) (1993) 199–220.

[30] C.A. Welty, An integrated representation for software development and discovery, Ph.D. Thesis, http://www.cs.vassar.edu/faculty/welty/papers/phd/ (Last visited: March 29, 2005).

[31] D.C. Berrios, A. Kehler, L.M. Fagan, Knowledge requirements for automated inference of medical textbook markup, Proc. AMIA Symp. (1999) 676–680.

[32] Y. Kalfoglou, M. Schorlemmer, IF-Map: an ontology-mapping method based on information flow theory, J. Data Semantics I, LNCS 2800, Springer, 2003, pp. 98–127.

[33] F. Wiesman, N. Roos, P. Vogt, Automatic ontology mapping for agent communication, in: Proceedings of the First International Joint Conference on Autonomous Agents, Multiagent, Systems, Bologna, Italy, 2002, pp. 563–564.

[34] N. Guarino, Formal ontology and information systems, in: N. Guarino (Ed.), Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, June 6–8, IOS Press, Amsterdam, 1998, pp. 3–15.

[35] B. Smith, Beyond concepts, or: ontology as reality representation, in: A. Varzi, L. Vieu (Eds.), Formal Ontology and Information Systems. Proceedings of the Third International Conference (FOIS 2004), IOS Press, Amsterdam, 2004, pp. 73–84.

[36] E.J. Lowe, A Survey of Metaphysics, Oxford University Press, Oxford, 2002.

Werner Ceusters is professor of health sciences in the department of psychiatry and coordinator of biomedical informatics at the University of Buffalo. He leads the Referent Tracking Unit in the Ontology Research Group of the NY State Center of Excellence in Bioinformatics and Life Sciences. He has additional degrees in informatics and knowledge engineering. He started a series of international research projects in medical natural language processing under the Third, Fourth and Fifth Research Frameworks of the European Commission through his R&D company Office Line Engineering nv. Since then, he has also been active in standardisation bodies related to medical terminology such as CEN/TC251/WG2. His current research is in the application of realist ontologies to electronic health records.

Barry Smith is Julian Park Distinguished professor of philosophy in the University at Buffalo (New York, US) and Director of the Institute for Formal Ontology and Medical Information Science in Saarbrücken, Germany. He studied at Oxford and Manchester, and has held faculty positions in Sheffield, Manchester, Liechtenstein and Leipzig, as well as visiting positions in Erlangen, Graz, Paris, Turku, Innsbruck, Padua, Hamburg, Konstanz, Malta, Leiden, Vienna and Koblenz. He is the author of some 400 scientific publications, including 15 authored or edited books, and editor of *The Monist: An International Quarterly Journal of General Philosophical Inquiry*. His research has been funded by the US, Swiss and Austrian National Science Foundations, the Volkswagen Foundation, and the European Union. In 2002, he received in recognition of his scientific achievements the €2.2 million Wolfgang Paul Award of the Alexander von Humboldt Foundation. Smith's current research focus is ontology and its applications in biomedicine and biomedical informatics, where he is working on a variety of projects relating to biomedical terminologies and electronic health records. He is also collaborating with Hernando de Soto, Director of the Institute for Liberty and Democracy in Lima, Peru, on the ontology of property rights and social development.