Proceedings of ICBO 2009

Towards an Ontological Representation of Resistance: The Case of MRSa

Albert Goldfain[1], Lindsay G. Cowell [2], Barry Smith[3]
[1]Blue Highway, Syracuse, New York;
[2] Duke University Medical Center, Durham, North Carolina;
[3] University at Buffalo, Buffalo, New York

Abstract

*This paper addresses a family of issues surrounding the biological phenomenon of resistance and its representation in realist ontologies. Resistance terms from various existing ontologies are examined and found to be either overly narrow, inconsistent, or otherwise problematic. We propose a more coherent ontological representation using the antibiotic resistance in Methicillin-Resistant Staphylococcus aureus (*MRSa*) as a case study.*

Introduction: IDO, SaIDO, and MRSa

The phenomenon of resistance is an important feature of biological reality, encompassing phenomena such as the resistance of an individual to specific diseases, the resistance of disorders to specific treatments, and the resistance of certain pathogens to certain drugs. As such, resistance is a phenomenon that needs to be captured in biomedical ontologies.

The Infectious Disease Ontology (IDO) consortium is developing a set of interoperable ontologies that together are intended to provide coverage of the infectious disease domain. At the core of the set is IDO itself, which provides a representation of all of these types of entities, drawn from both the biomedical and the clinical domains that are relevant to infectious diseases in general. Domain-specific extensions (e.g., pathogen-specific extensions) of this core IDO complete the set by providing ontology coverage of entities relevant to specific sub-domains of the infectious disease field. IDO is itself an extension of the Basic Formal Ontology (BFO).

The *Staphylococcus aureus* Infectious Disease Ontology (SaIDO) is an extension of IDO concerning *Staph aureus* (Sa) infection. Sa can be partitioned into two subtypes: Methicillin-Susceptible Sa (MSSa) and Methicillin-Resistant Sa (MRSa). The latter subtype is a defined class that is distinguished by its resistance to methicillin (and other β-lactam antibiotics). Due to its rapid evolution in the face of antibiotic selective pressures, MRSa has become the paradigm of resistance (a so-called "superbug"), and has drawn significant attention from NIAID/NIH, CDC, and biomedical researchers throughout the developed world.

Subtypes of Sa can also be specified by assigning bacterial strains to clonal complexes based on genotypic differences. Variants can differ in their degree of resistance and in the types of drug to which they are resistant, forming a continuum, in terms of which Sa can be (and is) categorized. This provides one powerful reason to produce an ontologically correct representation of resistance.

In this communication, we consider the issues arising from the representation of resistance in realist ontologies and specifically, in IDO. We will focus our attention on the antibiotic resistance of MRSa to methicillin as a case-study.

Ontological Issues Stemming from Resistance

An important principle for realist ontology development is to avoid as far as possible the use of negative differentia (e.g., 'nonphysical', 'not part of the heart') in formulating definitions. This "positivity design principle" enforces the use of terms which capture information about the entities represented in the ontology rather than information about the state of our knowledge at some given time.[1]

At some level, however, resistance seems to require a negative aspect for its description. After all, a continuant is resistant precisely when something does *not* happen. John's resistance to marriage entails a host of processes that do *not* happen (for example, John does not buy an engagement ring, does not get a marriage license, and so forth). In the case of MRSa, resistance to methicillin entails that a process of cell wall formation is *not* interfered with. The key is that the implicit negativity of resistance is only a semantic feature of the description *at some level.* The biologi-

cal phenomenon of resistance is manifested at various levels of biological reality: genes, cells and their parts, organisms, and populations. Negative descriptions at a macro-scale here mask the positive and active aspects of resistance at the micro-scale. A comprehensive ontological treatment must, accordingly consider resistance at different levels of granularity.

In BFO-based ontologies, the **lacks** relation can be used to capture negative findings at one scale of biological description while avoiding the problems of using negative predicates or characteristics.[2] In describing resistance, we will have a need to say that an independent continuant does not exhibit a dependent continuant. As we will see below, this amounts to an independent continuant lacking a certain disposition.

Resistance is referred to by several disciplines: epidemiologists describe the spread of resistance in a population, the medical community speaks of patient resistance to disease and pathogen resistance to drugs, and geneticists make reference to the genes that confer resistance when certain alleles are present. The IDO suite of ontologies must capture all of these discipline-specific aspects of resistance and the relations between them.

Resistance in Existing Ontologies

We surveyed the treatment of resistance in existing ontologies.

Gene Ontology (GO). A general treatment of resistance is outside the scope of the GO, as resistance is not a biological process, molecular function, or cellular component. Within the sub-ontology of biological processes, however, GO contains the term 'response to drug', with synonyms 'drug resistance' and 'drug susceptibility/resistance'.

> [GO:0042493] Response to Drug: A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a drug stimulus.

This treatment is inadequate because the narrower term "drug resistance" is made a synonym of the broader term "response to drug". Resistance arises spontaneously as the result of genetic diversification. The presence of the drug provides a fitness advantage to those cells or viral particles that have the resistance conferring gene or mutation, thus they outcompete the susceptible individuals. The resistance is not a direct response to the drug stimulus, although the manifestation of resistance may be a consequence of exposure to the drug. A response to a drug is a *process*, whereas resistance is a *continuant*. This error (although GO usually is very good at preventing this confusion), arises from an inadequate analysis of resistance. Finally, the GO definition of drug resistance seems to hinge on a 'change in state', but cells which do not change state are manifesting a 'response to a drug' just as much as those which do.

NCI Thesaurus. The NCI Thesaurus has the following entry for 'resistance':

> [C19391] Resistance: Natural or acquired mechanisms, functions, activities, or processes exhibited by an organism to maintain immunity to, or to resist the effects of, an antagonistic agent, e.g., pathogenic microorganism, toxin, drug.

The primary problems with this treatment of resistance are that: i) the definition is circular, since it uses 'resist' in defining 'resistance', and ii) the term 'resistance' is a child of "resistance process", making resistance a process, as in the GO, and excluding many types of resistance, because the definition of 'resistance process' is biased towards multicellular organism resistance mediated by host defense mechanisms.

SNOMED-CT. SNOMED-CT contains the entry 'drug resistance (disorder)' with two defining relationships:

> Drug Resistance Is a Drug-Related Disorder
> Drug Resistance has Causative Agent (Attribute) Drug or Medicament.

With a parent term like 'drug-related disorder', it is clear that this definition is given from the perspective of the patient. From the perspective of a pathogen (*qua* organism) or tumor, for example, drug resistance is not a disorder, but rather a benefit. Also, the definition specifies that drug resistance is caused by a drug, but resistance is caused typically by the presence of a gene or mutation. It is only the manifestation of resistance that results from the presence of the drug. Finally, as with other terms in SNOMED, only necessary but not sufficient conditions for drug resistance are provided. Good definitions should spell out both.

Infectious Disease Ontology (IDO). IDO includes the term 'protective resistance', the definition of which attempts to address some of these problems:

> Protective resistance is a disposition that inheres in an organism by virtue of the fact that the organism has a part (e.g., a gene product), the disposition of which is to ensure a physiologic response of a certain degree

> to a potentially damaging entity *P*, or to prevent the completion of some process caused by *P*, thereby protecting the organism from or mitigating the damaging effects of *P*.

In the next section, we describe the ontological case study that helped lead us to this definition.

Towards a More Robust Ontological Treatment

To better understand the representational demands posed by resistance (and to expose the problems raised by this and similar phenomena from an ontological point of view), it will be useful to go through a detailed example. We choose drug resistance for a single combination of pathogen, antibiotic, and resistance-mechanism types. In this section we will sketch the outlines of a formal representation of the resistance of MRSa to methicillin as conferred by PBP2a, a penicillin binding protein (PBP) and a product of the gene *mecA*. Both methicillin and penicillin are β-lactam antibiotics and, for the purposes of our formalization, a PBP can be considered to be a *methicillin* binding protein. Chambers[3] gives a concise description of this form of resistance: "[M]ethicillin resistance in staphylococci is due to expression of PBP2a, a novel, low-affinity PBP for which there is no homologue in methicillin-susceptible strains". We formalize this information as a set of triples expressing the relevant ontological relationships. We also include a series of inference rules that would lead a logic-driven reasoner to deduce from the triples that MRSa is resistant to methicillin. Alongside the statistical techniques employed in biology, it will one day be desirable for automated reasoners to compute antibiotic resistance from logical formalizations. Using ontologies as predictive tools will guide treatment decisions and support automated drug discovery.

The terms used in our representation will be derived from IDO, GO, and the Protein Ontology. The relations used are drawn from the OBO Relation Ontology (RO) and its extensions (see http://www.obofoundry.org/ro/). Naïvely, we could introduce a new relation **resistant_to** and represent the entire situation as MRSa **resistant_to** methicillin. However, this would hide the complexity of the mechanisms of resistance working at a smaller scale and elminate many important inferences about resistance. Also, it is important to avoid a proliferation of relations in the OBO Foundry, since restriction to a small set of relations promotes reuse and interoperability of the constituent ontologies.

A more faithful representation requires at least the following components (where **is_a** and **has_part** are used for relations between both continuant and occurrent universals):

[1] bacterium **is_a** organism
[2] MRSa **is_a** bacterium
[3] synthesis_of_peptidoglycan **is_a** process and **has_participant** Penicillin_Binding_Protein (PBP)
[4] PBP **has_function_realized_as_process** synthesis_of_peptidoglycan
[5] Bacterial_cell_wall **is_location_of** PBP
[6] Canonically, synthesis_of_peptidoglycan **results_in_development_of** bacterial_cell_wall
[7] formation_of_bacterial_cell_wall **is_a** process
[8] PBP2a **is_a** PBP
[9] methicillin_PBP_binding_process **is_a** binding process that **has_participant**s methicillin and PBP
[10] affinity_to_methicillin **disposition_of** some PBP to undergo a methicillin_PBP_binding_process that is **realized** in the presence of a methicillin.
[11] methicillin_PBP_binding_process **negatively_regulates** synthesis_of_peptidoglycan.
[12] PBP2a **lacks** affinity_to_methicillin
[13] *mecA* **is_a** gene
[14] MRSa **has_part** *mecA*
[15] *mecA* **generically_specifies** PBP2a_production
[16] PBP2a_production **results_in_formation_of** PBP2a

These triples will be used along with several rules of inference and derived facts (labeled *IRn* and D*n* respectively in what follows). For readability, all variables are italicized and initial universal quantifier symbols are suppressed. First, we specify that **is_a** and **has_part** (for both continuants and occurrents) are transitive, allowing us to derive some basic taxonomic facts about the domain:

(IR1) *x* **is_a** *y* & *y* **is_a** *z* → *x* **is_a** *z*
(IR2) *x* **has_part** *y* & *y* **has_part** *z* → *x* **has_part** *z*
(D1) MRSa **is_a** organism

The parts of an organism are the products of the organism's expressed genes, and these products are located in the appropriate places:

(IR3) (*o* **is_a** organism & *g* **is_a** gene & *o* **has_part** *g* &
g **generically_specifies** *proc* &
proc **results_in_formation_of** *prod* &
o **has_part** *locp* & *locp* **is_location_of** *prod*) →
o **has_part** *prod* **located_in** *locp*
(D2) MRSa **has_part** PBP2a **located_in** bacterial_cell_wall

The inference rule (IR3) makes a few simplifying assumptions. Since not all genes are expressed, we are only modeling the situation in which *g* is an expressed gene. We also assume that the process *proc* leading to *prod* is active, and that the single gene *g* generically specifies *proc* (rather than a set of genes).

If a continuant lacks a disposition to undergo a process in some situation, and that process negatively regulates a second process which has the continuant

as a participant, then the continuant participates in the second process in that situation:

(IR4) *p* **lacks** disposition to undergo *proc1* **realized** in situation *s*
& *proc1* **negatively_regulates** *proc2*
& *proc2* **has_participant** *p* →
In situation *s*, *p* **participates_in** *proc2*
(D3) In the presence of methicillin, PBP2a **participates_in** synthesis_of_peptidoglycan.

This lack of a disposition (i.e., the affinity to methicillin) has a categorical basis in the fact that methicillin binds to PBPs and prevents them from carrying out their function. However, PBP2a lacks this affinity, so the presence of methicillin does not prevent the essential sub-processes of cell-wall construction in MRSa.

If an organism has a continuant as a part and that part participates in a process in some situation, then the process unfolds in the organism in that situation. Finally, if a process unfolds in an organism in some situation and the process results in the development of a continuant which (canonically) is a part of the organism, then the organism has the continuant as a part in that situation.

(IR5) In situation s, *p1* **participates_in** *proc* & *p1* **located_in** *p2*
& o has_part p2 → *proc* **unfolds_in** *o* in situation *s*.
(D4) synthesis_of_peptidoglycan **unfolds_in** MRSa in the presence of methicillin.
(IR6) In situation *s*, *proc* **unfolds_in** *o* &
Canonically, *proc* **results_in_development_of** *p* →
p **part_of** *o* in situation *s*
(D5) Bacterial_cell_wall **part_of** MRSa in the presence of methicillin.

The canonical cell wall is a rigid configuration of peptidoglycan. From the perspective of MRSa, the canonical cell wall is a healthy one. The assertion (D5) captures the active, and thus positive, microphysical side of the resistance coin.

However the chain of reasoning here presents a puzzle. What does the lack of a disposition in (IR4) amount to? Consider the following pair:

(A) Continuant C lacks disposition D to undergo process P in situation S
(B) Continuant C undergoes P in a situation of type S.

Both (A) and (B) can be true at the same time. In fact the conjunction of (A) and (B) implies that (B) happens for a non-dispositional reason (i.e., (B) is not, in the corresponding case, a manifestation of the disposition D). Even if John lacks the disposition to feel hungry when in the presence of sushi, he may still feel hungry in such a situation because he has been fasting for three days. We need a way to say that PBP2a *necessarily* lacks affinity to methicillin in order to permit the relevant cell-wall formation.

Mereological Issues

If we take resistance to be a specifically dependant continuant that inheres in an independent continuant, then we must still answer some mereological questions: Is the resistance of PBP2a (i.e., of a part) identical to the resistance of the cell (i.e., of the including whole)? Furthermore, is cell resistance identical to the resistance of a portion of tissue in which the cell resides or the containing host organism or, for that matter, of the containing population? The ontology of resistance must address which scales of biological reality resistant continuants occupy, and the identity of resistance across scales.

Another issue that should be addressed at different scales of biological reality is the way in which facts at each scale are used to *explain* the phenomenon of resistance. At the genetic scale, MRSa having *mecA* and MSSa lacking *mecA* are explanatory. At the cellular level (D5) is explanatory.

Conclusion

We have seen that resistance is an important multi-scale and multi-domain phenomenon, often with a one-to-many relationship between a resistant organism and the underlying mechanisms of resistance. Several desiderata for an ontological representation were found lacking in existing ontologies. Our preliminary formalization of resistance honors both a *positivity design principle* and a *principle of non-proliferation of relations*, both of which are sound principles for the design of effective ontologies. Some puzzles remain (e.g., an account for the lack of a disposition), but further study of resistance will have great benefits for biomedical ontologies.

Acknowledgements

This work was funded by the National Institutes of Health through Grant R01 AI 77706-01. Smith's contributions were also funded through the NIH Roadmap for Medical Research, Grant 1 U 54 HG004028 (National Center for Biomedical Ontology).

References

1. Spear A. "Ontology for the twenty-first century: An introduction with recommendations". 2006. http://www.ifomis.org/bfo/manual.pdf
2. Ceusters W, Elkin P, and Smith B. "Negative findings in electronic health records and biomedical ontologies: A realist approach". *Int J Med Inform*. 2007;76(3):s326—s333.

3. Chambers H. “Penicillin-binding protein—mediated resistance in pneumococci and staphylococci”. *J of Infectious Disease*. 1999; 179(2):S353—359.