

BY ALBERT GOLDFAIN, PhD



## Canonicity and Disease Ontologies

Ontologies provide biomedical researchers with an inventory of the universal features of reality across organisms, biomedical disciplines, and levels of granularity. In capturing what is universal, there is often a need to refer to what is prototypical, representative, true-by-default, and statistically expected. In other words, we often need a reference for *canonical* entities and relationships. Clinically speaking, canonical facts for human beings

and other forms of departure from the norm. Treatment decisions are made on the basis of how (and to what extent) a patient deviates from a canonical life plan. Thus, a reasoner needs a coherent representation for the canonical that is compatible with biomedical ontologies.

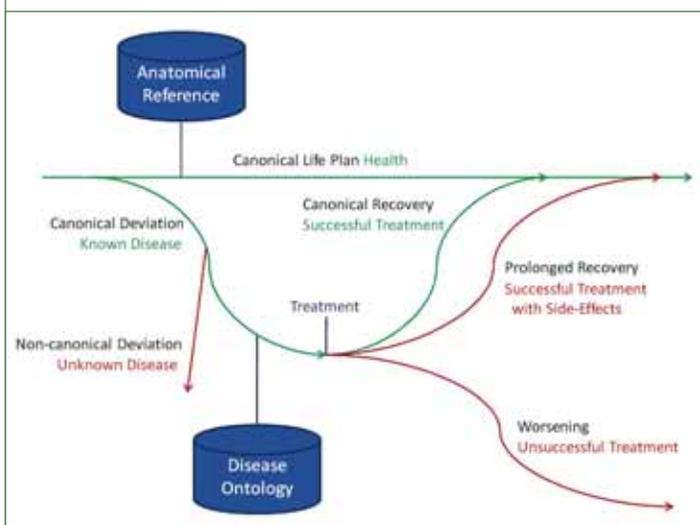
If ontologies are to be used in service of *computational* reasoning (e.g., in clinical decision support), the notion of canonicity must be formalized and given a precise semantics. Some work has been

done along these lines for anatomical ontologies [1], in which the range of quantified variables in logical formulae is restricted to anatomical entities. Such ontologies capture information about properly ordered parts undergoing proper functions, manifesting healthy dispositions, and playing proper roles. This approach can also be extended to disease ontologies, which capture disordered parts along with

references computationally amounts to using a similarity metric in a space of clinical measurements. Ontologies must include definitions for what these measurements are measurements of.

If I break my arm in a way that has never been clinically observed; or after treatment, my arm fracture worsens, does not heal in the expected amount of time, or does not heal at all, these would be *non-canonical* deviations from health, treatment, and recovery. Such instances can then be compared with a canonical reference regarding the clinical expectation. Non-canonical deviations, then, can serve as important signposts of unknown diseases or disorders, unsuccessful treatments, or erroneous outlier data.

Without a formal description of the canonical, a computational reasoner can only compare instance data to more instance data. This may be fine if the instances are sufficiently large in number and are drawn from a sufficiently representative population, but this is often an idealized situation. A reasoner who is given a dataset of human arms fractured playing football will not compute a sufficiently general prototype of a fractured arm because the instances are more likely to be the same sorts of high impact fractures. Such a reasoner would ignore the existing general knowledge about different types of arm fractures and the relations between them. Canonical entities serve as a compact summarization of general knowledge. They enhance ontologies by providing a baseline from which a deviation can be logically described, quantified, and measured.



include: body temperature is 98.6 degrees Fahrenheit, pregnancy lasts nine months, and adults have 32 teeth. These are context-independent clinical expectations that are the result of broad scientific consensus. Ontologically speaking, there are no instances of canonical humans with canonical parts functioning in canonical ways. Nevertheless, the canonical representation serves as a useful and economical starting point for describing and understanding disease

their associated malfunctioning, and the dispositions for disease that they give rise to. Clinical treatment is only possible because there are known patterns in how disorders become the physical basis for disease—essentially, *canonical* deviations from health. It is this sort of information that is covered by ontologies of disease and disorder. So, for example, my right arm currently resembles the human right arm as described in an anatomical reference. If I break my arm tomorrow, it may resemble a canonically fractured arm (if the fracture is of a known type). We cannot say that my arm is an instance of either the canonical human arm today or the canonical fractured arm tomorrow, but these reference points are essential for reasoning about my arm (as an instance of the universal arm) and its change of state. Comparing my arm to either of these

### DETAILS

Albert Goldfain, PhD, is a researcher for Blue Highway, LLC. [www.blue-highway.com]. He is currently working as a postdoctoral associate on the Infectious Disease Ontology [www.infectiousdiseaseontology.org].

### REFERENCES

- [1] Neuhaus, Fabian and Barry Smith (2007), “Modeling Principles and Methodologies—Relations in Anatomical Ontologies”, in Albert Burger, Duncan Davidson, and Richard Baldock (Eds) *Anatomy Ontologies for Bioinformatics: Principles and Practice* (Springer: New York), p. 289-306. □