Maps of Medical Reason: Applying Knowledge Graphs and Artificial Intelligence in Medical Education and Practice

Bill Cope, Mary Kalantzis, ChengXiang Zhai, Andrea Krussel, Duane Searsmith, Duncan Ferguson, Richard Tapping, and Yerko Berrocal

1 Addressing Biodigital Convergence in Theory and Practice

'Biodigital convergence and its various implications, from biotechnology to bioeconomy, seem to promise a Copernican shift in our current way of life,' say the editors of this volume (Peters et al. 2021a: 6). The biodigital is 'an emerging configuration' (Peters et al. 2021b). The editors also speak to the notion of a 'postdigital' age, taking up an idea first contemplated by Nicholas Negroponte (1998): 'Its literal form, [digital] technology, is already beginning to be taken for granted'. Soon, 'like air and

B. Cope (✉)
College of Education and College of Engineering, University of Illinois, Champaign, IL, USA
e-mail: billcope@illinois.edu

M. Kalantzis
College of Education, University of Illinois, Champaign, IL, USA
e-mail: marykalantzis@illinois.edu

C. Zhai · D. Searsmith
College of Engineering, University of Illinois, Champaign, IL, USA
e-mail: czhai@illinois.edu; dsearsmi@illinois.edu

A. Krussel
Washington University School of Medicine, St. Louis, MO, USA
e-mail: krussela@wustl.edu

D. Ferguson
College of Veterinary Medicine, University of Illinois, Champaign, IL, USA
e-mail: dcf@illinois.edu

R. Tapping · Y. Berrocal
University of Illinois College of Medicine, Peoria, Peoria, IL, USA
e-mail: berrocal@uic.edu

M. A. Peters et al. (eds.), *Bioinformational Philosophy and Postdigital Knowledge Ecologies*, Postdigital Science and Education,
https://doi.org/10.1007/978-3-030-95006-4_8

drinking water, being digital will be noticed only by its absence, not its presence… Computers will be a sweeping yet invisible part of our everyday lives: We'll live in them, wear them, even eat them. A computer a day will keep the doctor away.'

If we are to take 'postdigital' at its word, we have become so thoroughly 'in' the digital that we have become 'post' by virtue of its barely noticeable ubiquity. However, like other 'posts', we need to suspect sweeping assertions of rupture, as if for instance the 'postmodern' were not a variant of the modern, and the 'posthuman' not a variant of the human. The digital, after all, is not so new, beginning perhaps with Claude Shannon's insight that electrical switches could represent two numbers—on and off; zero and one; which in turn could represent Boolean logic (Shannon 1938). In any event, underlying algorithmic processes are binary, and at most only partly and secondarily digital. And we can never be 'post' to the extent that the meanings can only be partially represented and calculated with quantifying media—bodily meanings and medical conditions, for instance. Our Fitbits do not erase the difference between body and its calculability. The bio and the digital do not become each other. The digital and the bio are fundamentally and irreducibly different from each other. The digital is no more than a quantitative, conceptual representation of limited features of the bio. Hence, Jandrić et al. (2018: 895) conclude: 'The postdigital is hard to define; messy; unpredictable; digital and analog; technological and non-technological; biological and informational. The postdigital is both a rupture in our existing theories and their continuation.'

Inspired by the themes of this volume, the chapter that follows speaks both philosophically and technically, and theoretically as well as practically. Philosophically, we want to speak to theoretical limits as well as the enormous potentials of algorithmic reason. Practically and technically, we want to describe a project in which we have been developing knowledge graph software to support medical education and electronic health records. This project has been supported by a series of three grants from Jump Applied Research through Community Health through Engineering and Simulation (ARCHES) program, a partnership between Jump Simulation and Education Center at OSF HealthCare and the Health Care Engineering Systems Center in the Grainger College of Engineering at the University of Illinois.

Our thesis is as follows: the biodigital is today both ubiquitous, as evidenced in rapidly expanding use of computable information in modern medicine, and an oxymoron, because the bio and the digital are irreducibly different things. As a consequence, our focus is the relation between these two.

Elsewhere, two of the research team have developed a metaontology centered around the idea of 'transposition' or movements in meaning appliable to digitally-mediated representation and communication (Cope and Kalantzis 2020; Kalantzis and Cope 2020). Here we have identified four transpositions between the qualities of biolife and the quantities in which they may be calculable (Cope et al. 2020). These are as follows.

Namability: the character collocations of medical labels can be tracked for similarity by search and other forms of text mining—the terms in a medical ontology, for instance—but computers can have no conception of what these labels mean.

Countability: instances of labelled things can be counted, but counting is always limited by the criterial features of a label, losing in the process irreducible materiality of the singular referrent and the subtleties of its context. This is an intrinsic limitation if population-based medical science, whatever its undoubted benefits.

Measurability: clines of progression can be measured, often now with automated sensors (temperatures, heart rates, pharmacology etc.), but such numbers are mostly limited to linear vectors.

Renderability: digital capture tools can render medical images on a two-dimensional plane, and 3D printing can construct three-dimensional medical models and even prosthetic body parts, but the materiality of rendered image and object is always fundamentally different from its referent.

These four things, and no more, can be performed by the mechanics of the digital. These are what the digital offers us, sometimes a miraculous servant for the cure of our bodies, but these transpositions also set its absolute limits. The limits of the digital are the limits of the transposability of qualitative biomeanings into quantity.

2 Ontologies, in Philosophy and Medical Practice

Theoretically, this chapter works at two levels, applying the concept 'ontology' philosophically at one level, and technically at another. Philosophically, the term 'ontology' connects material being with the immaterial of its understanding, our viscerally experienced lives and bodies connected with models or schemas of specialized medical understanding. In traditional philosophical terms, ontology is a relation between the material (bodies) and the ideal (medical reason). Technically, in the digital era ontologies perform the same function, as schemas that represent and model the material world. In our case, medical ontologies to describe bodies and their ailments. They transpose the ideal and the material through a backwards-and-forwards play between conceptualization and practice**.** In our case, the medical understandings captured in medical ontologies are connected in medical practice with the biomedical materiality of bodies. One of many such connections is that between bodies in a qualitative experiential sense and their quantitative calculability. This is the extent of 'biodigital convergence'.

On the question of calculability, the relation of the ideal of the calculability with the material of the world to which it refers, Gottfried Leibniz led the modern charge. Here he is, writing in 1666:

> I have found an astonishing thing, which is, that we can represent all sorts of truths and consequences by Numbers … [T]here are certain primitive Terms which can be posited if not absolutely, at least relatively to us, and then all the results of reasoning can be determined in numerical fashion, and even with respect to those forms of reasoning in which the given data do not suffice for an absolute answer to the question, we could still determine mathematically the degree of probability… and where there are disputes among persons, we can simply say, Let us calculate, without further ado, in order to see who is right. (Leibniz 1951: 50–51)

Chris Anderson, editor of Wired magazine, was still saying a version of this in 2008.

> Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology … This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear … With enough data, the numbers speak for themselves. (Anderson 2008)

We disagree, for the reasons that we lay out in this chapter. Of course, we live in an era of 'big data', of massively networked and integrated information. But purely algorithmic reason only works by leveraging human reason. One such leveraging is the process of inference in statistical text mining (Zhai and Massung 2016). Text consists of stings of characters which in combination reference things and ideas. Computers can never be intelligent in the sense of understanding ideas, but they can point to repetitions in strings of characters that come together into words that may make sense to humans. Statistical work with text can only point to latent semantics (Landauer et al. 2007). After the patterns of textual characters are counted, the limits of this process are the limits of natural language with its baffling ambiguities, infinitely variable context dependencies, and polysemy that even the best of dictionaries find hard to disentangle. Supervised machine learning helps, where humans supplement numerical or textual meaning with labels that are meaningful to them. This helps the machine detect future such patterns; and unsupervised machine learning where the machine finds statistical clusters or outliers that may warrant labelling. In both cases, the thinking can never be smarter than the labels that humans apply.

So, on top of the laborious but otherwise dumb calculation of frequencies of alphanumeric characters, the textual labels are crucial. Folksonomy is commonsense, ad hoc and frequently spontaneous labelling, of which hashtags have become a prominent example in the era of social media (Guy and Tonkin 2006). Taxonomy (hierarchical) and ontology (multidimensional) have become crucial tools for representation of meaning in the digital era.

In their basic functions, taxonomy and ontology are as old as the human mind—beginning with the extraordinarily complex totemic, kinship, and natural classification systems of First Peoples (Kalantzis and Cope 2020: 42–47). For modernity, Immanuel Kant proposed a 'categorical imperative', or the process by which mind applies its categories of meaning to the material world encountered by the human body. In our utilitarian twenty-first century, digitized ontologies are pervasive. Kant could be proud of the rigorous intent of the categorical systems insisted upon by the medical insurance companies in order to argue the cost of medical interventions.

An example: Intelligent Medical Objects (IMO) is a company in the Research Park at the University of Illinois that teases out the nuances of how human bodies work, or don't so well in the case of illness and death. 'Object' is a revealing word, because the manner of relating of concepts is digital ontologies are not merely conceptual; they purport reference objects of the material world, in this case manifest physiological conditions. These are not mere ideas or figures of cognition. They are 'objects', arranged in some sort of order, and this order is a series of determinate relations. Though not a commercial concern for the IMO company, this is where the philosophical notion of ontology meets the technical one. The medical labels dwell in the realm of the ideal, but they oriented to the realm of the material, people's bodies.

'Clinicians are asked to see 6–10 patients per hour & do all the documentation', says the PowerPoint presentation that is used to sell the IMO software product. 'The most expensive resource in the healthcare ecosystem is currently being used to do the bulk of documentation via the Electronic Health Record (EHR). How do you extract maximal value for your investment?'

IMO offers a standardized classification scheme by means of which medical vendors can share electronic records about a patient's medical condition, preserving 'the truth of clinical intent'. Not only is there a problem of accurate coding. There are two main classification schemes, ICD (the International Statistical Classification of Diseases and Related Health Problems[1]) and SNOMED (Systematized Nomenclature of Medicine[2]). ICD exists in a succession of versions, older records in ICD-9, and newer records in ICD-10 and after that, ICD-11. IMO provides apps, accessible on computers and phones, for looking up the terminology associated with different medical conditions across ICD-9, ICD-10, ICD-11, SNOMED, and other specialized medical ontologies.

IMO also analyzes synonyms that emerge in medical practice and maps these to the standard ontologies. For instance, IMO has a term 'abnormal excitement', which maps to ICD version 9 code 799.29, 'other symptoms involving emotional state'. Version 10 of ICD codes this R45.0, 'nervousness'. SNOMED codes it 247006004, 'Over-excitement', or 'Uncontrollable excitement'. At this point, medical classifications begin to run into another life-defining ontology, the Diagnostic and Statistical Manual of Mental Disorders.[3]

The historians of ICD trace the origins of formal classification of medical conditions to registrations of the causes of death from Italy in the mid-fifteenth century and England in the mid-sixteenth. ICD had its beginnings in the International Statistical Congress, which first met in Brussels in 1853, when the 'CD' part of the acronym stood for 'causes of death'. At the 1860 meeting in Paris, Florence Nightingale used death classification statistics to show the causes of hospital deaths and how they could be reduced (Bostridge 2008: 11–12).

The first version of the International List of Causes of Death was adopted at the Chicago congress of 1893 (Moriyama et al. 2011: 11–12). Since then, ICD has gone through eleven major versions, including digitization in the third quarter of the twentieth century. The title 'Causes of Death' was changed to 'Classification of Diseases' in 1949 when the World Health Organization took responsibility for it. WHO now hosts periodical revision conferences and manages the revision process. ICD-10 was released in 1994, ICD-11 in 2018. ICD-11 expands the number of

[1] See https://www.who.int/standards/classifications/classification-of-diseases. Accessed 1 October 2021.

[2] See https://www.snomed.org/. Accessed 1 October 2021.

[3] See https://www.psychiatry.org/psychiatrists/practice/dsm. Accessed 1 October 2021.

codes available to 55,000, up from the 14,400 in ICD-10. Meanwhile, SNOMED, in development since 1965 and now controlled by a London-based not-for-profit, was created to describe a range of pathologies and clinical processes—311,000 in total—not all of which are adequately captured in ICD.

Nobody could conceivably remember or be able to speak more than a few of 55,000 or 311,000 things, which is why IMO created the look-up app as a textual prosthesis for medical professionals. ICD-10 is a carefully ordered classification scheme, divided into in sections that are marked auspiciously with roman numerals. It has sections on various bodily systems: IX 'circulatory', X 'respiratory', XI 'digestive', IV 'endocrine'. Then there are some strange system conjunctions: XIII the 'musculoskeletal system', and XIV 'genitourinary system'—muscles are very different from bones and reproduction very different from urinating, but in the body, these things work together or are near each other.

Ontologies connect the objects of their reference in a number of kinds of relation. For instance, one thing may be a kind of another, or it may be an instance of another. 'Malignant neoplasm of the breast' is a kind of 'neoplasm'. This is type of coherence can be visualized taxonomically: one thing is an instance of concept that happens more than once, such as a single case of breast cancer. More broadly encompassing concepts can group subsets of narrower concepts. Ontologies can also represent things that are parts of another. For example, 'Sprain and strain of ankle' involves joints and ligaments, constituent parts of ankles. They can represent patterns of action in chains of cause and effect. 'Whooping cough due to Bordetella pertussis' is different by dint of its cause from 'Whooping cough due to Bordetella parapertussis'.

The coherence of ontologies is also just as much a matter of relations of difference, where as much importance is afforded to: not-a-kind of; not-a-part-of; does not have certain properties; or does not cause. In medicine, differential diagnosis is the process of distinguishing the 'is' from the 'is-not', even though the symptoms may have created initial uncertainty. Medicine then becomes a practice of weighing evidence in order to make categorical judgment. Differentials in medicine are close but nevertheless important differences. And there are differences that are just irrelevancies, or informational 'noise'.

No two relations are the same. We can use 'kind of', 'part of', 'property of', and 'cause of' as rough heuristics for relations. But muscles do not connect with bones in the same way that the environmental conditions of bodies and brains connect to emotional states.

This irreducible specificity of relations is also the reason we need big lists of things, codified and the subject of general agreement in the digital era. As well as describing the body and its ailments, GeoNames[4] is for places, Ethnologue[5] for languages, Chemical Markup Language[6] for chemistry, product numbering and

[4] See http://www.geonames.org/. Accessed 1 October 2021.

[5] See https://www.ethnologue.com/. Accessed 1 October 2021.

[6] See https://www.xml-cml.org/. Accessed 1 October 2021.

classification systems for purchasable things.[7] The list of lists in digital modernity is long, covering billions of the most useful and important things that can be meant, many of which, as it happens, may impact medical conditions. More meaning is to be found in the unique configurations of relations in these lists than can be found in purely algorithmic or logical work of so-called ‘artificial intelligence’.

Of course, even the most fastidiously organized ontologies are rough and indeterminate in places. In the case of ICD, some are internal to bodies, some external, but nevertheless objects of very different orders: organs or parts of the anatomy such as VIII ‘eyes’, VII ‘ears’, XII ‘skin’; acquired conditions such as I ‘infectious diseases’ and II ‘neoplasms’ (cancers); external effects such as XIX ‘injuries’; conditions that may have been inherited in the form of XVII ‘congenital malformations’; conditions that may not even be medical, at least in their origins, but which might now be classified as V ‘mental or behavioral disorders’; and stuff that happens in XXI ‘contact with health systems’. Cross-classification clarifications are offered in the form of inclusions and exclusions. XVI ‘perinatal conditions’, we are told, includes conditions whose origins are in pregnancy even though the baby dies later, but they exclude congenital malformations.

For all its agonizing order, and after a century and a half of institutional agonizing about its ordering, ICD still has the appearance of a ramshackle list. This is not because our medical thinking is flawed, but because the material world is endlessly varied and complex. The list is as ramshackle as the particularities and relations of biolife itself. And it is as fallible as the politics of the construction of the medical self, where old maladies such as homosexuality are no longer that, and new conditions appear, such as post-traumatic stress disorder. ‘Classification systems’, say Bowker and Star (2000: 61), ‘simultaneously represent the world “out there,” the organizational context of their application and the political and social roots of that context. Many of these concepts are matters of judgment and thus contention. ‘Excitement’, ‘nervousness’, and ‘emotional states’ traverse vast territories of human experience, and the point at which these become a medical condition as distinct from healthy life may at times become a contentious matter between patients, doctors, and insurance companies.

Then there is the frequently-appearing but nevertheless disquieting notion of ‘other’—‘other infectious diseases’, ‘unspecified mental disorder’, ‘neoplasms of uncertain or unknown behavior’, ‘other ill-defined and unspecified causes of mortality’, ‘provisional assignment of new diseases of uncertain etiology or emergency use’, to mention just a few labels for uncertainty in ICD-10. If an ontology is to encompass all possibilities in a domain, it has to countenance as-yet or in-the-moment unknown possibilities.

In the professional domain of medicine, the historically evolved ontology of ICD speaks an un-natural language. For all the open-ended possibilities of medical science, the aim of an ontology is to reference the world—in this case, the human body

[7] See https://www.gs1.org/standards/barcodes/ean-upc. Accessed 1 October 2021.

and its maladies—in ways that are more precise than the natural language of everyday or vernacular experience of sickness and health.

How, then, do digitally represented labels arrayed in conceptual schemas connect with the bodies they are labelling? At this point, we come to rely on the philosophical notion of ontology. This is our twenty-first century replacement for metaphysics, tracing the relations between the material and the ideal. The material of ontology is the immanent if complex order in the world, in nature, society and history. The ideal of ontology is the conceptualization of these meanings, not only by means of the artifact of language, but with other tools of representation including image, space and embodied feeling. The ideal is integrally connected with the material, but the one is never a straightforward reflection of the other. The ideal can exceed the material, in conjecture and imagination, in the service of a medical diagnosis or scientific hypothesis, for instance. And the material exceed the ideal, for example in the as-yet undiscovered but eventually knowable (Cope and Kalantzis 2020: 280–303).

Returning now to the specifics of medical ontologies. The labels are ideal in the sense that they are ideational constructs, and the bodies to which they refer are material. The two are of course connected. There is no label in a medical ontology without a material reference to which it can point in the body, brain, or behavioral manifestations of mind—either retrospectively or potentially. There is no meaningful material body without our making sense of it. But the ideal and the material are not mirror reflections of each other. The material can exceed the ideal—things that are as yet unexplained, for instance, in a particular case or in medical science. And the ideal can exceed the material—diagnoses about the causes of illness, or prognoses about progression and the effects of treatment. Ontology is a complex dialectic, a play between the ideal and the material (Cope and Kalantzis 2020: 302–10).

3 Knowledge Graphs, in Theory and Practice

Digital ontologies can be represented visually as concept maps or knowledge graphs. The concepts of an ontology can be represented as labelled containers, joined by lines to indicate one kind of relation or another: something that is a part of something else; something that is a kind of something else; something that has a namable property; or something that is a cause of something else. Medical ontologies can be represented taxonomically, as tree diagrams showing parent/child and sibling relations. Knowledge graphs, however, allow endless complexity where each node can be represented in multiple relations with other nodes, and different kinds of relations between nodes can be specified as labels for the arrows that connect them. To agree on labels and relations for a domain like medicine, as we have seen in the case of the history of ICD, is a long historical and open-ended social process.

The emergence of the social web has led to a shift in meaning making in the direction of more collective endeavors, of which ICD online is an example. This process can be found in present-day representation of large-scale collections of knowledge with emphasis on aggregation and integration, utilizing standardized schemes for scalability, structure, identity, relationship, and provenance. This phenomenon is clearly evidenced by the rise of large-scale knowledge graphs and their growing importance to sense making at scale.

The knowledge graph is not a novel development in knowledge representation or automated reasoning. In early AI research expert systems employed various mathematical logics and rule-based reasoners to represent knowledge and perform inference—with emphasis on capturing human expertise to approximate human level reasoning. Knowledge was thought of as 'descriptions, relationships, and procedures' (Hayes-Roth 1983). Semantic networks were used as visual representations of first order predicate logic (FOPL) where nodes represented entities and links were thought of as predicate relationships (Dietterich and Michalski 1983). Reasoning was performed by applying logic rules over typed entities to generate new entity relationships and attributes.

The current day graph database systems evolved gradually, in response to improvements in hardware performance and storage of ever-increasing amounts of data—including curated stores, and evolution of mature standards of data identity, semantics, and federation. Outside of academia, prior to the 1980s, data was stored in customized hierarchical structures largely influenced by hardware limitations and with a premium placed on efficiency of insertion and retrieval.

The 1980s through the end of the century saw the development of relational database technologies, based on the relational algebra. These technologies also emphasized data operational efficiencies over the leveraging the power of the data itself (i.e., relations between data elements were limited), with tables of columnar data types joined to other tables via key fields and queried using the Structured Query Language (SQL). Triple stores (RDF) and No-SQL technologies developed soon after which allowed data of any shape to be stored in key value systems, and relational knowledge of data entities could be represented at a finer grain. Graph database technologies in general use began to emerge around 2010 and have been improving over the last decade. The graph database systems of today allow relations as first-class objects with reasoning engines that fully realize the knowledge graph at scale. In general terms, Gosnell and Broecheler characterize the historical evolution as database from hierarchical models, to relational, to NoSQL, and in the 2020s, graph thinking (Fig. 1).

The widespread currency of the contemporary knowledge graph can in large part be traced Google's application of the technology to its search. Launched in 2012, Knowledge Graph presents summary text and images about a search topic in a panel on the right side of the screen. A search on the musician 'Taj Mahal' works because Knowledge Graph knows the difference between the singer, the tomb in Agra, and Donald Trump's failed casino in Atlantic City. Google knows enough about you from your search history and the history of searchers like you to decide which Taj Mahal you are most likely to be interested in. This is because the string of characters 'Taj Mahal' has been classified into a number of different meanings across multiple ontologies.

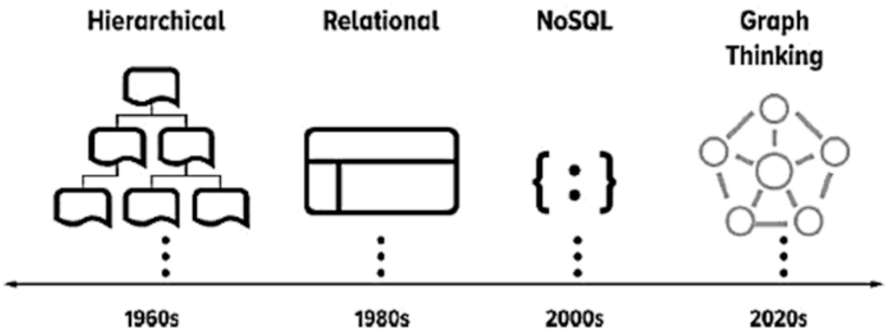

Fig. 1 Evolution of database technologies (Gosnell and Broecheler 2020: 3)

'Things, not strings', has become the Google Knowledge Graph mantra (Singhal 2012). Behind Knowledge Graph is a repository, Knowledge Vault. Unlike previous, text-based extraction, which can be very noisy, Knowledge Vault 'combines extractions from Web content… with prior knowledge derived from existing knowledge repositories' (Dong et al. 2014). Such knowledge repositories have been hand-curated and over a long period of time—medical ontologies, for instance. Google is secretive about its Knowledge Graph technology (Paulheim 2017: 2), but one suspects that its power lies in Knowledge Vault, where the copying of a huge amount of web content, including previously hand-curated ontologies, is at least as powerful and probably more powerful than data mining of unstructured text.

'Defined abstractly', says Kerjriwal (2019), 'a knowledge graph is a graph-theoretic representation of human knowledge such that it can be ingested with semantics by a machine'. However, graph conceptualizations of the world are infinite, capable of generating too much information to the point where they are practically unreadable. Each graph needs to be selective, a small and germane collocation of nodes and specified relations—whether these are machine suggestions, or machine-readable human suggestions (Cañas et al. 2015). These may function pedagogically or to support thinking, as advance concept organizers (Ausubel 2000), concept maps (Novak 2010), or learning progressions (Shi et al. 2020).

Gutiérrez and Sequeda argue that today, data represents a commodity, tied to bits and formats, devoid of any meaning in and of itself. Knowledge, on the other hand, has been thought of as a 'paradigmatic immaterial' object living only in people's minds and language. Over the decades computer scientists have gradually developed the techniques and systems to 'materially support knowledge'. The fusion of knowledge and data thus creates meaning, and knowledge graphs are a manifestation of this vision at scale (Gutiérrez and Sequeda 2021: 104).

The life sciences have been early adopters of semantic web technologies (Chen et al. 2013), and as such they are also heavily involved in advancing knowledge graph research and application. One challenge of knowledge graphs in the life sciences is the integration of disparate knowledge and data sources into one cohesive graph. The Covid-19 Community is a project funded by National Science Foundation

(NSF) to integrate 'environmental datasets to help researchers analyze the interplay between host, pathogen, the environment, and COVID 19' (Rose et al. 2020). The project is using the Neo4j graph database software to integrate disparate data sources to enable reasoning and search in the domain of Covid-19. Over eighteen such datasets are utilized spanning diverse domains such as geography (GeoNames),[4] Covid-19 data resources like the Covid-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University (Covid-19 Data Repository),[8] and general medical resources like the Disease Ontology (Disease Ontology[11]). The graph contains over ten million nodes and thirty-three million relationships. At a high level, these knowledge sources can be broken down into the following categories: metadata; biological data and literature; locations, epidemiological data; and population characteristics.

The network is updated nightly via a staged workflow. Public data sources are ingested (caching reduces redundant processing), and analytics process the data to produce node and relationship data which is integrated into the knowledge graph. To facilitate better integration, the project added their own metadata in the form of nodes and relationships. This is type information used to label data elements. For example, node types like 'strain', 'gene', and 'city' are used to provide type information and to link types via relationships in a graph wide manner. Data provenance was also added by including data source nodes that identify the origin of various data elements.

Another important feature for interoperability across various resources is the use of unique identifiers. The European Union has developed a resolution service called identifiers.org which maintains unique identifiers in the form of compact URIs for over 700 life science resources. This is of critical importance to integrate and link concepts from the various data sources. In other instances where no such service is available, custom identifiers were constructed using attributes of the data elements. Challenges that were encountered in constructing this resource included the sheer variety of file formats, data types, and modalities of data access. Additionally, the sheer volume of the data to be processed while it is changing in real-time was a challenge for synchronization and machine resources. Data inconsistencies also require strategies for bridging gaps, detecting file format errors and the like. There was also a need for domain knowledge to bridge gaps in knowledge representations across domains and at differing scales.

In the following sections we describe the work we have been doing to apply medical ontologies to medical education and health records.

[8] See https://www.openicpsr.org/openicpsr/covid19. Accessed 1 October 2021.

4 Medical Informatics: Where Algorithmic Reason Meets the Semantics of the Body

The research and development we have been undertaking addresses some methodological questions for evidence-based medical science that go to the very heart of its information systems. Leading computer scientist, Judea Pearl, questions the limits of purely statistical and correlational approaches to data, recommending they should be supplemented by 'causal models [that] facilitate the evaluation of the effect of novel actions… that were unanticipated during the construction of the model' (Pearl 2009: 307). He makes a distinction between an actual cause (in a single event) and general cause in populations (type-level) (309–310). His conclusion is that 'to achieve human level intelligence, learning machines need the guidance of a model of reality' (Pearl 2018). In the medical domain, ontologies offer such models of reality. The statistical processes of artificial intelligence depend on the precision of the labels applied to data and used to model reality.

In a single medical case, millions, even potentially billions of highly specified concepts might be available to model causal reality. In the specialized medical domain, there are a number of widely used models, some with long and storied histories. Some we have already mentioned in this chapter, others we mention now to expand this picture: The International Statistical Classification of Diseases and Related Health Problems (ICD);[1] the International Classification of Functioning, Disability and Health (ICF);[9] the Systematized Nomenclature of Medicine (SNOMED);[2] Logical Observation Identifiers Names and Codes (LOINC);[10] and The Drug Ontology (DRON).[11] Then there are ontologies of everyday non-medical things capable of identifying, with a high degree of precision, contextual variables that may be highly relevant to a clinical case, including place (GeoNames),[4] time and event (iCal),[12] demographic profile (age, gender, race/ethnicity etc.), occupational classification (SIC: Standard Occupational Classification),[13] or objects in the form of identifiable products (IAN: International Article Number).[7] Transmission of an infectious disease, for instance, might be identified at the conjunction of a precisely specified place, event, occupation, demographic, or product.

The key characteristic of these ontologies—their millions of classifiers and the billions of data points classified in standardized schemas with unique identifiers—is their semantic precision. Their model of reality is much more finely specified than natural language, with far less potential ambiguity. They can also be arrayed in taxonomic or knowledge graph structures where term-to-term relations (e.g. parent/

[9] See https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health. Accessed 1 October 2021.

[10] See https://loinc.org/. Accessed 1 October 2021.

[11] See https://www.ebi.ac.uk/ols/ontologies/dron. Accessed 1 October 2021.

[12] See https://icalendar.org/. Accessed 1 October 2021.

[13] See https://www.bls.gov/soc/. Accessed 1 October 2021.

child, sibling, and other relations) offer insights into the basic components and relationships of a given model of reality (Cope et al. 2011; Sowa 2000).

Purely algorithmic, text mining approaches to artificial intelligence rely to a large degree on statistical analysis of natural language where meaning is principally limited to character collocations in 'stemmed' words (Brown et al. 1991; Kalantzis and Cope 2020). Meanings are subsequently inferred in processes of 'latent semantic analysis' (Landauer et al. 2007). As powerful as these analyses have become, they can be all-the-more powerful when AI is applied to processes of semantic modeling.

The main questions for our project are, how can ontologies deepen processes of artificial intelligence, and how can machine learning generated from human interaction with knowledge graphs add detail and depth to those graphs? In prior research, Zhai and colleagues have used the ConceptNet knowledge base to add supplementary semantics to text mining (Kotov and Zhai 2012). They have leveraged public domain knowledge graphs to improve text-based prediction (Jiang et al. 2018). In the broad area of bioinformatics, they have developed a schema of entity relation semantics for insects for a project researching the genetics of bees (He et al. 2010). The lessons that we have learned from these previous studies are that ontologies can enhance artificial intelligence in multiple ways including facilitating human-in-the-loop collaboration with AI, improving explainability, and addressing the challenge of data sparseness in supervised machine learning. All these benefits are especially important in medical informatics and can be realized via developing innovative technologies for creating, updating, and maintaining comprehensive domain-specific ontologies.

The final two sections of this chapter address the specific development objectives and application testing processes in our current projects.

5 Application Project 1: Medical Education

In a first phase of our current research, we have web-based developed knowledge graph software called CGMap (Common Ground Map), applicable across a number of domains.[14] The medical application of this software we have called MedMap. In the current phase, we are applying and testing this software in two domains: mapping of clinical cases in medical education (MedMap Application 1), and electronic health records (MedMap Application 2).

In the MedMap clinical case application, the web-based environment that we have been developing has been designed to contribute to medical education by supporting holistic, critical, and problem-based learning. In a review of the literature on critical clinical thinking, Benner, Hughes and Sutphen adopt a definition of critical thinking as 'purposeful, self-regulatory judgment that uses cognitive tools such as

[14] See https://newlearningonline.com/cgscholar/projects/medlang. Accessed 1 October 2021.

interpretation, analysis, evaluation, inference, and explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations on which judgment is based'. Applying this to the context of medical practice, they point out that 'the growing body of research, patient acuity, and complexity of care demand higher-order thinking skills. Critical thinking involves the application of knowledge and experience to identify patient problems and to direct clinical judgments and actions that result in positive patient outcomes.' As they argue,

> clinicians and medical scientists alike need multiple thinking strategies, such as critical thinking, clinical judgment, diagnostic reasoning, deliberative rationality, scientific reasoning, dialogue, argument, and creative thinking, and so on. In particular, clinicians need forethought and an ongoing grasp of a patients health status and care needs trajectory, which requires an assessment of their own clarity and understanding of the situation at hand, critical reflection, critical reasoning, and clinical judgment. (Benner et al. 2008)

Mukherjee (2015) describes his training as a physician in these terms: 'The profusion of facts obscured a deeper and more significant problem: the reconciliation between knowledge (certain, fixed, perfect, concrete) and clinical wisdom (uncertain, fluid, imperfect, abstract)'. Gambrill (2012) notes that evidence-based clinical practice is rooted in a willingness to recognize the intrinsic uncertainty of clinical decision-making. Research by Bordage and colleagues demonstrate that critical clinical thinking requires the analysis of systematic semantic patterns (Bordage and Lemieux 1991; Bordage 2007). Failing to teach clinicians about clinical uncertainty has been referred to as 'the greatest deficiency of medical education' (Djulbegovic 2004). Evidence-based medicine (EBM) requires the skillset to develop answerable questions relevant to a case, and to answer these questions with an honest and open appraisal of research findings (Braddock et al. 1999). Learning by doing is emphasized in EBM and evaluation of clinical cases provides such practice.

Individual or exemplary case analysis is a key aspect of clinical problem-solving and medical learning. If used early and with appropriate scaffolding, it has the potential to demonstrate the relevance of the 'pre-clinical' subjects within a medical curriculum. Clinical case study can serve as an important supplement to the acquisition of medical content knowledge by learners. The societal need we are attempting to address is the education of medical professionals who make sound judgments based on clinical evidence, using a variety of human and data sources to make these judgments. Such habits should be encouraged as early as possible and not left to less structured fast-paced clinical training periods.

For the purposes of medical education and training, the individual cases analyzed by students may be hypothetical, created by an instructor or instructional designer, or real cases found in the literature of medical science. The disciplinary challenge is differential diagnosis, or to develop the capacity to distinguish a particular medical condition from others that may have similar features, and to apply and monitor appropriate therapy.

The challenge for evidence-based medicine is that symptomatic and contextual information is necessarily limited, and incomplete and multiple conditions may be present. As a consequence, diagnosis is a matter of professional judgment. Furthermore, paradigms of medical knowledge are changing, where the focus is not only the generic individual whose physiology is assumed to be universally replicable, but where enormous variation is also recognized based on finely determined environmental context, genetic profile, and demographic variables (Wilson and Cleary 1995). Greenhalgh et al. (2014) address what they consider to be a crisis of evidence-based medicine where ‘contemporary healthcare’s complex economic, political, technological and commercial context has tended to steer the evidence-based agenda towards populations, statistics, risk, and spurious certainty’. As a counterbalance they recommend the exercise of ‘judgment not rules’ in relation to individual patients, and, in medical education, the development of ‘clinical skills, understanding and applying research evidence, and reflecting and deliberating about complex cases’.

This occurs through a process that can formally be characterized as argumentation, involving hypothesis, claims supported by evidence, rebuttal of potential counter-claims, and preliminary professional judgment (Cope et al. 2013; Gillies and Khan 2009; Toulmin 2003; van Eemeren et al. 2002). Typical media for clinical case analysis are case textbooks (Geha and Notarangelo 2016), project-based learning (Greeno 1998), team-based learning (Michaelsen and Sweet 2008), and oral-discussion of cases in situ in medical contexts. Each of these media has its limitations: textbooks tend to transmit knowledge more than encourage active problem solving and professional collaboration (Boulos et al. 2006); project-based learning is difficult and expensive to assess given the complexity of the artifacts created, and often the lack of explicit clarity in their documentation (Kreijns et al. 2003); and in-situ oral engagements are ephemeral where most interactions are invisible to the instructor, and leaving few analyzable traces of the participants’ thinking processes (Artino et al. 2014).

MedMap, is a web-browser ontology-suggestion, diagramming, and visualization tool that offers a way to document the features of a medical case using tags from widely used medical ontologies. It supports students as they determine the connections between these features such as possible causal relations or to create a decision tree to plan treatment. MedMap leverages medical and other contextually relevant ontologies, suggesting labels with a high degree of semantic precision. It adds semantic awareness to the electronic text of a case analysis, by: (1) making *coding suggestions* that add a layer of formal semantics to the clinical case analysis; (2) offering a *concept visualization* or mapping tool for deeper analysis of the underlying medical logic of a clinical case; (3) providing opportunities for peers, self, and instructors to code annotations with formalized medical vocabulary, thereby also training the system using machine learning methods.

User mappings are automatically analyzed, contributing to the knowledge embodied in the system via mechanisms of supervised machine learning. The system also offers conjectures about possible labels, in processes of unsupervised or semi-supervised machine learning. In computer science, this project supplements the algorithmic power of machine and deep learning with under-exploited areas of data modelling, formalized ontologies, and rigorously defined and structured semantics. MedMap’s user interface takes the form of a concept map visualization (Novak 2010; Olmanson et al. 2016; Schroeder et al. 2018; Tergan 2005; Villalon and Calvo 2011). In Fig. 2, the left side of the case creator's screen is a multimodal documentation space. When the user highlights elements of text or media on the left, the system suggests possible matches for terms specified in one or more ontologies in a smart suggestion system. If selected by the creator, a node will appear in the right panel of the screen. The creator can then begin a concept map connecting nodes according to standard medical relations. This builds a second, semantically formalized, diagrammatic layer of meaning into the case documentation, or what we have called 'explain protocols', an extension of the 'think-aloud' protocols (Ericsson and Simon 1993).

Machine learning and artificial intelligence processes are validated in a collaborative peer review process by: (1) peer reviewers, and (2) by the case creator when they accept a peer reviewer's recommendation. Once validated, the ontology will have been trained to a higher level of machine intelligence to make suggestions for subsequent cases. By these means, users add a rich multidimensionality to the essentially two-dimensional taxonomic structures of legacy ontologies.

The project emphasizes problem-based learning and critical clinical thinking, supporting the documentation of hypothetical or actual clinical cases with tagging

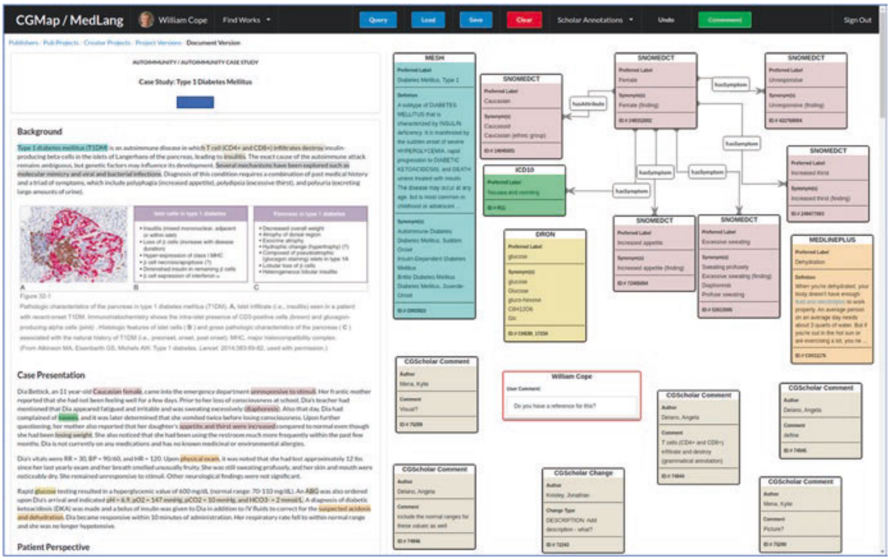

Fig. 2 Clinical case study report, written up on the left side of the screen by the student, and on the right side, a knowledge graph created to explain the condition and provide diagnosis. Left side: Multimodal case documentation with color-coded annotations according to ontology items in the visualization. Each color represents a different ontology. Right side: A medical logic model visualization, where each node and relation is linked to standard medical ontologies: e.g. International Statistical Classification of Diseases (ICD); the International Classification of Functioning, Disability and Health (ICF); the Systematized Nomenclature of Medicine (SNOMED); Logical Observation Identifiers Names and Codes (LOINC); and The Drug Ontology (DRON). Nodes are identified in an easy look-up and suggestion tool—highlighting a part of the data on the left, generates node suggestions for the visualization on the right

from structured medical vocabularies. The objective of this project is to develop in medical students critical, holistic and evidence-based thinking. In support of these objectives, MedMap makes suggestions based on standard medical and other ontologies. MedMap calls for explicit clinical reasoning mapped in a standards, ontology-based case visualization. By leveraging user interactions and inputs, the system captures new knowledge and improve system performance in a synergistic machine-user relationship. For instance: when users connect terms to be found in multiple ontologies or synonyms; when they annotate a patient-friendly vernacular term with a concept from an ontology; or when a user connects a related term not immediately suggested by the taxonomic relationships already present within the ontology. In other words, users by their concept selection and medical logic mappings become involved in training the system, validating its semantics, and thereby, progressively develop its semantic intelligence and the value of its suggestions to subsequent student or researcher users.

In supervised machine learning, users identify semantically precise nodes and draw visualizations of clinical logic models. In unsupervised machine learning, the system offers label suggestions and node connections based on the semantic structure of the supporting ontologies and the previous actions of users. Suggestions vary depending upon the focus of case analysis (e.g. establishing diagnosis), its reasoning (e.g. hypotheses, pathophysiology, evidence, etc.) or its treatment (e.g. linkage between clinical problems and therapeutic choices).

We have been testing the online, cloud-based software module with students and faculty at the University of Illinois College of Medicine, Peoria, and graduate students at Washington University School of Medicine in St. Louis. Students are being presented with cases framed in a clinical case documentation. An online peer review process applies and extends medical knowledge through collaborative clinical analysis: case documentation > peer review > revision based on feedback > publication to the student's e-portfolio. Key aspects of this innovation and testing include:

1. A web-based *multimodal clinical case documentation space*, where data, image, video, 3D imaging, ambiently collected case data, and other media offering supporting empirical evidence, are embedded within the case narrative.
2. A *lookup system* where, in support of the process of documentation, the medical case creator can identify concepts from standardized medical and other ontologies. The case creator highlights the media item or text and connects with the concept.
3. A *suggestion system that* 'reads' the case during the process of documentation, suggesting on-the-fly annotations from medical and other ontologies, as well as connections that the machine has gleaned from other users' case documentation via machine learning processes.
4. The creator then maps a *case logic model* in the form of a *concept visualization*. They can also write unstructured comments to support their thinking processes, or leave open questions for consideration during later stages in the documentation process or by peers during the review process. (The white boxes in Fig. 2 are self-annotations by the case creator.)

5. The case is then submitted to *multiple peers for review*, who in a new version suggest changes including different or additional markup and a revised visualization based on a different extended logic models. (The brown boxes in Fig. 2 are comment annotations by peer reviewers.)
6. Synthesizing feedback from peers, the case creator revises their case and updates their logic model visualization, ready for *publication* by the instructor to the student's e-portfolio and possibly also to the class online community.

6 Application Project 2: Electronic Health Records

It is widely agreed that the health industry is destined to be revolutionized by AI (Reddy et al. 2019). Evidence-based medicine is increasingly data-centric, involving a symbiotic relationship between humans and machines. AI can play an important role in detecting patterns in individual cases, as well as detecting variation across multiple cases. However, Himmelstein et al. (2010) show that while hospital computing increases overall costs, it only modestly improves process measures of quality. A great deal of work still needs to be done to develop user-facing medical informatics systems that attain impacts for patient care promised by AI.

In the era of 'big data' and progressively more comprehensive systems for developing and sharing medical records, variability by case is increasingly visible, and indeed essential for precision medicine. Greenhalgh et al. (2014) address what they consider to be a crisis of evidence-based medicine where 'contemporary healthcare's complex economic, political, technological and commercial context has tended to steer the evidence-based agenda towards populations, statistics, risk, and spurious certainty'. As a counter-balance they recommend the exercise of 'judgment not rules in relation to individual patients, and the development of 'clinical skills, understanding and applying research evidence, and reflecting and deliberating about complex cases'.

Moreover, paradigms of medical knowledge are changing, where the focus is not only the generic individual whose physiology is assumed to be universally replicable, but where enormous variation is also recognized based on finely determined life history, genetic profile, environmental context, and demographic variables (Wilson and Cleary 1995). The relevance of MedMap solution lies in: (1) the precision of documentation of cases with the support of standards-based and ontology-originated medical and everyday concepts; (2) tying these concepts together into a medical logic model in support of clinical reasoning and tracing the thinking underlying decision pathways.

A key question then for our ongoing research is to create a more precise record of the particularities of case-to-case differences. To address this question, MedMap is a web-browser semantic suggestion, tagging, and visualization tool that documents the features of a medical case with precision. It leverages medical and everyday life ontologies, and the data they contain, suggesting labels with a high degree of semantic precision for supervised machine learning, and offers conjectures about possible labels in the case of unsupervised machine learning. MedMap supports research scientists, students, trainees, practitioners, and data analysts as they make connections between these features such as possible causal relations or the creation of decision trees to plan treatment.

The prototype we have been developing aligns with the Precision Medicine Initiative, a US government effort involving the National Institutes of Health, aimed at leveraging new opportunities in access, aggregation, and analysis of health data. This initiative has two major foci, both of which are addressed in this project: (1) to develop personalized approaches to patient care; (2) to support researchers to make new discoveries. Our particular interest in 'precision' is the detection of semantic configurations in single or small numbers of cases. To give the example of a novel pathogen, this may begin in a localized context and a unique configuration of events. Then, in the first critical stages of transmission, there were only a few cases. Once the pathogen spreads across a wider population, the course of the disease may present differently from patient to patient, depending on a unique configuration in of medical history, genetic propensities, environment, and other factors. While sharing detectable and statistically measurable features, every subsequent case is unique, and every medical outcome causally related to a unique configuration of circumstances. A key question for this work is to put a more precise record of the particularities of case-to-case differences into practice.

Although the focus of this research is single cases or small numbers of cases, and the varied presentation of subsequent cases, the semantic processes we are advocating also enrich big data and AI across populations at any scale by making data points smaller, more precise, and their relations specified with less ambiguity. Cross-mapping between ontologies yields further precision.

This project addresses four specific areas of need in AI-supported medical informatics in medical case documentation:

1. Efficient and effective medical communication and error-free collaboration, where medical records accurately represent case histories and provide useful information to a range of specialist team members. The Intelligent Medical Record tool *offers suggestions* informed by standardized medical and other ontologies as well as AI comparisons with similar, de-identified records.
2. Explicit *clinical reasoning* and decision tree mapped in a standards (ontology)-based case visualization or map.
3. Sharing of deidentified data in the service of public health and precision medicine. The system we are proposing will offer a lightweight yet standards-based *mechanism for interoperability* across different professional and personal health record and data systems.
4. Early *ontology and AI-supported detection of a single case or small numbers* of cases, and variation in the course of a clinical condition across cases.

The focus of this project is the detection of patterns of significance in a single case based on *intelligent suggestions* and the *mapping of clinical reasoning*. For the purposes of development of a prototype in this project, we are using the source

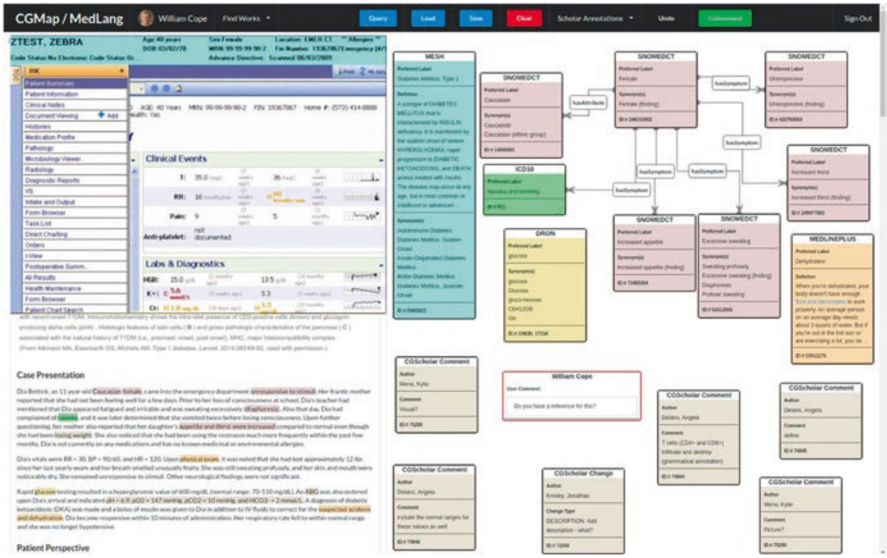

Fig. 3 Case annotation and clinical reasoning tool. The left side of the screen is the case material, for the purposes of illustration here consisting of a screenshot of an EPIC electronic medical record, followed by plain text in our proof-of concept tool. The case documented on the left has annotations color-coded according to ontology items in the visualization. Right side: annotations drawn from medical ontologies, linked by clinical reasoning and decision tree relations. White boxes are unstructured notes to self; brown boxes are unstructured notes from others

OpenEHR.[15] However, we are not suggesting that health providers should change their medical records systems. Rather, we are proposing an intelligent overlay over any medical records system in the form of either:

1. a *'private' web browser window* for data mirroring side-by-side with the existing system, or
2. a *shadow box* browser overlay for secure simultaneous double entry, or
3. *automatic export/import* between the current medical records and the proposed intelligent record.

We have attempted to minimize the effort required to learn and use the system and limit it to a single, browser- based screen with a simple drag-and-drop user interface, as illustrated in Fig. 3.

This particular project uses leverages MedMap to create an intelligent medical record. The left side of the screen features a medical record, illustrated above with a screenshot of an EPIC record and the right side is the case annotation and clinical reasoning tool. In this design, the medical practitioner highlights a term or label on the left, and the annotation tool suggests possible terms from multiple ontologies. The practitioner then builds a logic model or clinical reasoning visualization on the

[15] See https://www.openehr.org/. Accessed 1 October 2021.

right side of the screen. We are currently testing our hypothesis, which is that the proposed tool will improve: (1) diagnosis and decision support; (2) precision identification of clinical conditions; (3) anonymized data sharing across a distributed medical informatics system.

Interoperability and the ownership of medical record data have been identified as a key challenge for the future evolution of medical informatics and personalized medicine. With varying data models and overlapping ontologies, considerable challenges are raised for the exchange of data (Benson and Grieve 2016; Garde et al. 2007; McGuire 2018; Reis et al. 2017). A related challenge is the question of ownership of medical data (Zhu et al. 2012). In our proposed intelligent record, we address these problems by creating a connected but independent layer of data.

Addressing both these challenges, on 9 March 2020, the U.S. Department of Health and Human Services announced a requirement that both public and private organizations in the health industry should establish systems to share information, while positioning the patient so they are in control of that data.[16] This points to new openings and opportunities for the approach we are proposing now. The potential benefits of the Intelligent Medical Record system we are developing are:

1. *Standards-based suggestions*—supporting medical students, trainees, and professionals as they document a case in both classroom and clinic.
2. Case *logic model visualization*—making clinical reasoning and decision trees explicit to self and colleagues.
3. A *highly granular, standards-based markup* of the clinical case—supporting discovery and machine analysis across multiple cases.
4. A widely *distributed medical knowledge system*, where broad range of medical practitioners can contribute the development of medical datasets and scientific knowledge with the support of ontology- based knowledge scaffolds and explanatory logic models.
5. AI applications—Unsupervised machine learning processes will detect associations across multiple cases, identifying *patterns of similarity* which, although uniquely configured, may share features pointing to causality and potential implications across populations, small and large. By these mechanisms, ontologies can be supplemented and become 'smarter' the more they are used.

Together, these benefits offer patients more personalized medical care, with more accurate diagnosis and customized decision support. Through the development of MedMap, we aim to develop of a new generation of medical informatics and learning environments for medical students and practitioners.

In developing MedMap, the team is employing agile development strategies with short, focused release cycles. This approach is a lightweight and highly collaborative process that is well suited to research-based software development and lightly structured for diverse project teams (Stober and Hansmann 2009). This method

[16] See https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie. Accessed 1 October 2021.

emphasizes rapid, flexible evolution of software that is put into user hands at an early stage for frequent evaluation and feedback. This is being collected in highly controlled, small-group cognitive labs. At the end of each cycle a new functional evolution of the prototype is released, and the next development cycle build upon that work. In terms of software architecture, the approach of this project has been to develop three system layers:

1. A *metaontology* layer (Cope and Kalantzis 2020) encompassing all widely used medical and related ontologies and capable of creating an overarching framework for patient information the medical domain. For trials, we are applying MedMap to OpenEHR.
2. An *interlanguage* layer, to address schema synonyms, inter-schema overlaps and cross-schema connections (Cope et al. 2011).
3. A *permissions* layer, offering patients a range of options, including for instance: allowing identified data to be shared across medical providers with whom they might need to engage; and sharing de-identified data for the benefit of other patients, including in the case of this proposal, the intelligent record suggestion system. Patients could also grant permission for sharing their data from personal health apps and records, ambient intelligent sensors, and implanted devices (Patterson 2013; Roehrs et al. 2019). Permission to share geolocation data, for instance, might offer insights into the communication of pathogens—not unlike the aggregated data now used to offer traffic density reports in map apps. This would be accompanied by privacy and security guarantees for all users, including those who have not opted into data sharing (Dagher et al. 2018; Staudigel et al. 2017).

We are teaming with the Institute for Informatics at Washington University School of Medicine in St. Louis to test the feasibility of the MedMap overlay with residents and fellows from different specialties and subspecialties. Physicians and informaticians will test and use the software to document clinical cases. For this project, we will overlay Open EHR medical records. Following the trial, participants will fill out surveys and participate in interviews and focus group sessions to determine feasibility of this system. The software will then be revised based on trial performance and user feedback.

As a web application that supports case or project documentation that uses structured vocabularies for semantic parsing, author suggestions, formative and summative assessments, and semantically aware data mining and meta-analysis, we believe MedMap may have the potential to transform the ways in which clinicians and researchers interact with the electronic health record and patient data.

7 Reconciling the Bio and the Digital in Bioinformation

In this chapter we have attempted to explore the connections between the bio and the digital, because nowadays the digital serves our understanding of biolife and supports remedies for our bodily maladies.

Postdigital? The digital may slip from our consciousness as it ubiquitously inveigles itself into our lives, but it seems a perverse use of the Latin word for 'after' when the mechanics of the digital impacts our lives more and more.

Biodigital? Yes, to the degree that we increasingly use digital machines to master biolife. But this is hardly a singular reality when the bio and the digital are irreducibly different. The bio is material, where the digital entails a transposition of the meaning of the material into calculable quantity. However, as we have argued in this chapter, there are absolute limits to the transposability of the bio and digital meanings.

The focus of this chapter has been the explore the potentials of the bridge between the material of bodies and the ideal of their understanding by applying the philosophical and technical notion of 'ontology'.

Philosophically, this is a dialectic where, although the ideal and the material are integrally related, the ideal of conceptualization can sometimes exceed the immediately material (hence, modelling, conjecture, prognosis), and the material can exceed the immediately comprehensible (hence, uncertainty, discoverability, and diagnosis). Ontologies are models of material reality, and although the models are derived from material experience, in their immaterial ideality they are quite different. The ideal side of ontology is the paradigmatic, immaterial representation of reality, with degrees of latitude in the connection between the ideal and the material: hence uncertainty, absence, and the role of imagination in the discovery of the as-yet unknown. Such is the dialectic of knowledge. This ontology in the philosophical sense.

Technically, the ontologies of computer science model the world of experience for the purposes of systemic and scientifically-grounded labelling in digital information systems. Where the tables of databases were two-dimensional, knowledge graphs are multidimensional, and for this reason support complex semantics more effectively. Moreover, the limits of these systems are not merely digital; they are the limits of language, and beyond that the struggle of classification to make sense of the biomaterial world.

In this chapter, we have played out these ideas in the area of medical ontologies, arguing that purely algorithmic AI needs to be supplemented by human reasoning in the form of semantically-grounded data models.

Translating these ideas into practice, we have described two projects in which users represent ontologies visually and conceptually in the form of knowledge graphs. Our research questions are entirely practical. Can medical students' clinical reasoning be deepened by the use of knowledge graphs representing medical ontologies? And, can knowledge graphs support medical practitioners' thinking as a supplement to electronic health records? This is a work in progress, and we await empirical evidence.

References

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired, 16 July. https://www.wired.com/2008/06/pb-theory/. Accessed 1 October 2021.

Artino, A. R., Cleary, T. J., Dong, T., Hemmer, P. A., & Durning, S. J. (2014). Exploring Clinical Reasoning in Novices: A Self-regulated Learning Microanalytical Approach. *Medical Education, 48*(3), 280–291. https://doi.org/10.1111/medu.12303.

Ausubel, D. P. (2000). *The Acquisition and Retention of Knowledge: A Cognitive View*. Dordrecht: Springer.

Benner, P., Hughes, R. G., & Sutphen, M. (2008). Clinical Reasoning, Decisionmaking, and Action: Thinking Critically and Clinically. In R. G. Hughes (Ed.), *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*. Rockville, MD: Agency for Healthcare Research and Quality.

Benson, T., & Grieve, G. (2016). Why Interoperability Is Hard. In T. Benson & G. Grieve (Eds.), *Principles of Health Interoperability: SNOMED CT, HL7 and FHIR* (pp. 19–35). Cham: Springer. https://doi.org/10.1007/978-3-319-30370-3_2.

Bordage, G. (2007). Prototypes and Semantic Qualifiers: From Past to Present. *Medical Education, 41*(12), 1117–1121. https://doi.org/10.1111/j.1365-2923.2007.02919.x.

Bordage, G., & Lemieux, M. (1991). Semantic Structures and Diagnostic Thinking of Experts and Novices. *Academic Medicine, 66*(9), S70–S72. https://doi.org/10.1097/00001888-199109000-00045.

Bostridge, M. (2008). *Florence Nightingale: The Woman and Her Legend*. London: Viking.

Boulos, M. N., Maramba, K. I., & Wheeler, S. (2006). Wikis, Blogs and Podcasts: A New Generation of Web-based Tools for Virtual Collaborative Clinical Practice and Education. *BMC Medical Education, 6*, 41. https://doi.org/10.1186/1472-6920-6-41.

Bowker, G. C., & Star, S. L. (2000). Sorting Things Out: Classification and Its Consequences, Cambridge MA: MIT Press.

Braddock, C. H., Edwards, K. A., Hasenberg, N. M., Laidley, T. L., & Levinson, W. (1999). Informed Decision Making in Outpatient Practice. *Journal of the American Medical Association, 282*(24), 2313–2320. https://doi.org/10.1001/jama.282.24.2313.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1991). Word Sense Disambiguation Using Statistical Methods. In D. A. Appelt (Ed.), *ACL '91 Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics, Berkeley CA* (pp. 264–270). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.3115/981344.981378.

Cañas, A. J., Novak, J. D., & Reiska, P. (2015). How Good is My Concept Map? Am I a Good Cmapper? *Knowledge Management & E-Learning: An International Journal, 7*(1), 6–19. https://doi.org/10.34105/j.kmel.2015.07.002.

Chen, H., Yu, T., & Chen, J. Y. (2013). Semantic Web Meets Integrative Biology: A Survey. *Briefings in Bioinformatics, 14*(1), 109–125. https://doi.org/10.1093/bib/bbs014.

Cope, B., & Kalantzis, M. (2020). *Making Sense: Reference, Agency and Structure in a Grammar of Multimodal Meaning*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781316459645.

Cope, B., Kalantzis, M., & Magee, L. (2011). *Towards a Semantic Web: Connecting Knowledge in Academic Research*. Cambridge, UK: Elsevier.

Cope, B., Kalantzis, M., & Searsmith, D. (2020). Artificial Intelligence for Education: Knowledge and its Assessment in AI-enabled Learning Ecologies. *Educational Philosophy and Theory, 52*(16), 1–17. https://doi.org/10.1080/00131857.2020.1728732.

Cope, B., Kalantzis, M., Abd-El-Khalick, F., & Bagley, E. (2013). Science in Writing: Learning Scientific Argument in Principle and Practice. *E-Learning and Digital Media, 10*(4), 420–441. https://doi.org/10.2304/elea.2013.10.4.420.

Dagher, G. G., Mohler, J., Milojkovic, M., & Marella, P. B. (2018). Ancile: Privacy-preserving Framework for Access Control and Interoperability of Electronic Health Records Using

Blockchain Technology. *Sustainable Cities and Society, 39*, 283–297. https://doi.org/10.1016/j.scs.2018.02.014.

Dietterich, T. G., & Michalski, R. S. (1983). A Comparative Review of Selected Methods for Learning from Examples. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach* (pp. 51–60). Los Altos, CA: Morgan Kaufmann Publishers.

Djulbegovic, B. (2004). Lifting the Fog of Uncertainty from the Practice of Medicine. *British Medical Journal, 329*(7480), 1419–1420. https://dx.doi.org/10.1136%2Fbmj.329.7480.1419.

Dong, X. L. Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., & Zhang, W. (2014). Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. In J. Leskovec, W. Wang, & R. Ghani (Eds.), *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 601–610). New York: Association for Computing Machinery. https://doi.org/10.1145/2623330.2623623.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data.* Cambridge, MA: MIT Press.

Gambrill, E. (2012). Critical Thinking in Clinical Practice: Improving the Quality of Judgments and Decisions, Hoboken NJ: Wiley.

Garde, S., Knaup, P., Hovenga, E. J. S., & Heard, S. (2007). Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. *Methods of Information in Medicine, 46*(3), 323–343. https://doi.org/10.1160/me5001.

Geha, R., & Notarangelo, L. (2016). *Case Studies in Immunology: A Clinical Companion*. New York: Garland Science.

Gillies, R. M., & Khan, A. (2009). Promoting Reasoned Argumentation, Problem-solving and Learning During Small-group Work. *Cambridge Journal of Education, 39*(1), 7–27. https://doi.org/10.1080/03057640802701945.

Gosnell, D., & Broecheler, M. (2020). *The Practitioner's Guide to Graph Data*. Sebastopol, CA: O'Reilly Media.

Greenhalgh, T., Jeremy, H., & Neal M. (2014). Evidence Based Medicine: A Movement in Crisis, BMJ, 348(g3725).

Greeno, J. G. (1998). The Situativity of Knowing, Learning, and Research. *American Psychologist, 53*(1), 5–26. https://psycnet.apa.org/doi/10.1037/0003-066X.53.1.5.

Gutiérrez, C., & Sequeda, J. F. (2021). Knowledge Graphs: Tracking the Historical Events that Lead to the Interweaving of Data and Knowledge. *Communications of the ACM, 64*(3), 96–104. https://doi.org/10.1145/3418294.

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up Tags? *D-Lib, 12*(1). http://www.dlib.org/dlib/january06/guy/01guy.html. Accessed 1 October 2021.

Hayes-Roth, F. (1983). Using Proofs and Refutations to Learn from Experience. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach* (pp. 221-240). Berlin: Springer. https://doi.org/10.1007/978-3-662-12405-5_8.

He, X., Li, Y., Khetani, R., Sanders, B., Lu, Y., Ling, X., Zhai, C-X, & Schatz, B. (2010). BSQA: Integrated Text Mining Using Entity Relation Semantics Extracted from Biological Literature of Insects. *Nucleic Acids Research, 38*(2), 175–181. https://dx.doi.org/10.1093%2Fnar%2Fgkq544.

Himmelstein, D. U., Adam, W., & Steffie, W. (2010). Hospital Computing and the Costs and Quality of Care: A National Study, *The American Journal of Medicine, 123*(1), 40–46. https://doi.org/10.1016/j.amjmed.2009.09.004.

Jiang, S., Zhai, C-X., & Mei, Q. (2018). Exploiting Knowledge Graph to Improve Text-based Prediction. In N. Abe, H. Liu, C. Pu, X. Hu, N. Ahmed, M. Qiao, Y. Song, D. Kossmann, B. Liu, K. Lee, J. Tang, J. He, & J. Saltz (Eds.), *2018 IEEE International Conference on Big Data (Big Data)* (pp. 1407–1416). IEEE. https://doi.org/10.1109/BigData.2018.8622123.

Jandrić, P., Jeremy, K., Tina, B., Thomas, R., Juha, S., & Sarah, H. (2018). Postdigital Science and Education, *Educational Philosophy & Theory, 50*(10), 893–899. https://doi.org/10.1080/00131857.2018.1454000.

Kalantzis, M., & Cope, B. (2020). *Adding Sense: Context and Interest in a Grammar of Multimodal Meaning*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781108862059.

Kerjriwal, M. (2019). *Domain-Specific Knowledge Graph Construction*. Cham: Springer. https://doi.org/10.1007/978-3-030-12375-8.

Kotov, A., & Zhai, C-H. (2012). Tapping into Knowledge Base for Concept Feedback: Leveraging ConceptNet to Improve Search Results for Difficult Queries. In E. Agichtein & Y. Maarek (Eds.), *WSDM '12: Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 403–412). New York: Association for Computing Machinery. https://doi.org/10.1145/2124295.2124344.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the Pitfalls for Social Interaction in Computer-supported Collaborative Learning Environments: A Review of the Research. *Computers in Human Behavior, 19*(3), 335–353. https://doi.org/10.1016/S0747-5632(02)00057-2.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007*). Handbook of Latent Semantic Analysis*. New York: Routledge.

Leibniz, G. W. (1951). *Leibniz: Selections.* Trans. P. P. Weiner. New York: Charles Scribner's Sons.

McGuire, M. R. (2018). A Critique of Interoperability, Big Data, Artificial Intelligence and Medical Care in General Currently. *Health Science Journal, 12*(5), 1–2. https://doi.org/10.21767/1791-809X.1000592.

Michaelsen, L. K., & Sweet, M. (2008). The Essential Elements of Team-Based Learning. *New Directions for Teaching and Learning, 116*, 7–27. https://doi.org/10.1002/tl.330.

Moriyama, I. M., Loy, R. M., & Robb-Smith, A. H. T. (2011). *History of the Statistical Classification of Diseases and Causes of Death*. Hyattsville, MD: National Center for Health Statistics.

Mukherjee, S. (2015). *The Laws of Medicine: Field Notes from an Uncertain Science*. New York: Simon and Schuster.

Negroponte, N. (1998). Beyond Digital. Wired, December. https://web.media.mit.edu/~nicholas/Wired/WIRED6-12.html. Accessed 5 October 2021.

Novak, J. D. (2010). *Learning, Creating and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*. New York: Routledge.

Olmanson, J., Kennett, K., McCarthey, S. J., Searsmith, D., Cope, B., & Kalantzis, M. (2016). Visualizing Revision: Leveraging Student-Generated Between-Draft Diagramming Data in Support of Academic Writing Development. *Technology, Knowledge and Learning, 21*(1), 99–123. https://doi.org/10.1007/s10758-015-9265-5.

Patterson, H. (2013). Contextual Expectations of Privacy in Self-Generated Health Information Flows. Paper presented at TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy. https://doi.org/10.2139/ssrn.2242144.

Paulheim, H. (2017). Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. *Semantic Web, 8*(3), 489–508. https://doi.org/10.3233/SW-160218.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference.* Cambridge, UK: Cambridge University Press.

Pearl, J. (2018). Theoretical Impediments to Machine Learning with Seven Sparks from the Causal Revolution. arXiv:1801.04016.

Peters, M. A., Jandrić, P., & Hayes, S. (2021a). Biodigital Technologies and the Bioeconomy: The Global New Green Deal? *Educational Philosophy and Theory.* https://doi.org/10.1080/00131857.2020.1861938.

Peters, M. A., Jandrić, P., & Hayes, S. (2021b). Postdigital-Biodigital: An Emerging Configuration. *Educational Philosophy and Theory.* https://doi.org/10.1080/00131857.2020.1867108.

Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial Intelligence-enabled Healthcare Delivery. *Journal of the Royal Society of Medicine, 112*(1), 22–28. https://doi.org/10.1177/0141076818815510.

Reis, Z. S. N., Maia, T. A., Soriano Marcolino, M., Becerra-Posada, F., Novillo-Ortiz, D., & Pinho Ribeiro, A. L. (2017). Is There Evidence of Cost Benefits of Electronic Medical Records, Standards, or Interoperability in Hospital Information Systems? Overview of Systematic Reviews. *JMIR Med Inform, 5*(3), e26. https://doi.org/10.2196/medinform.7400.

Roehrs, A., da Costa, C. A., da Rosa Righi, R., Rigo, S. J., & Wichma, M. H. (2019). Toward a Model for Personal Health Record Interoperability. *IEEE Journal of Biomedical and Health Informatics, 23*(2), 867–73. https://doi.org/10.1109/jbhi.2018.2836138.

Rose, P. W., David, V., & Ilya, Z. (2020). COVID-19-Net: Integrating Health, Pathogen and Environmental Data into a Knowledge Graph for Case Tracking, Analysis, and Forecasting, https://github.com/covid-19-net/covid-19-community. Accessed 5 October 2021.

Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and Constructing Concept Maps: A Meta-analysis. *Educational Psychology Review, 30*, 431–55. https://doi.org/10.1007/s10648-017-9403-9.

Shannon, C. E. (1938). A Symbolic Analysis of Relay and Switching Circuits. *Transactions American Institute of Electrical Engineers, 57*, 471–495.

Shi, D., Wang, T., Xinga, H., & Xu, H. (2020). A Learning Path Recommendation Model Based on a Multidimensional Knowledge Graph Framework for e-Learning. *Knowledge-Based Systems*, 195, 105618. https://doi.org/10.1016/j.knosys.2020.105618.

Singhal, A. (2012). Introducing the Knowledge Graph: Things, Not Strings. Google blog, 16 May. https://www.blog.google/products/search/introducing-knowledge-graph-things-not/. Accessed 5 October 2021.

Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical and Computational Foundations.* Pacific Grove, CA: Brooks Cole.

Staudigel, M., Prokosch, H-U., & Kraus, S. (2017). An Abstraction Layer to Facilitate Technical Interoperability Between Medical Records and Knowledge Modules. In R. Röhrig, A. Timmer, & H. Binder (Eds.), *German Medical Data Sciences: : Visions and Bridges: Proceedings of the 62nd Annual Meeting of the German Association of Medical Informatics*. Oldenburg: German Association of Medical Informatics.

Stober, T., & Hansmann, U. (2009). *Agile Software Development: Best Practices for Large Software Development Projects*. Dordrecht: Springer.

Tergan, S-O. (2005). Digital Concept Maps for Managing Knowledge and Information. In S.-O. Tergan & T. Keller (Eds.), *Knowledge and Information Visualization* (pp. 185–204). Berlin: Springer.

Toulmin, S. (2003). *The Uses of Argument.* Cambridge, UK: Cambridge University Press.

van Eemeren, F. H., Grootendorst, R., & Snoek Henkemans, F. (2002). *Argumentation: Analysis, Evaluation, Presentation*. New York: Routledge.

Villalon, J., & Calvo, R. A. (2011). Concept Maps as Cognitive Visualizations of Writing Assignments. *Educational Technology and Society, 14*(3), 16–27.

Wilson, I. B., & Cleary, P. D. (1995). Linking Clinical Variables with Health-related Quality of Life: A Conceptual Model of Patient Outcomes. *JAMA, 273*(1), 59–65.

Zhai, C-H., & Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. Williston, VT: ACM and Morgan & Claypool.

Zhu, Q., Carl. G., & Tamer, B. (2012). Tragedy of Anticommons in Digital Right Management of Medical Records, in HealthSec '12. Bellevue WA.