Chapter 24
The Future of Ontology

Peter L. Elkin and Barry Smith

In the late 1800s, there began the age of the *manufacturing* economy. We built roads and highways and railroads, which allowed raw materials to be brought to factories and the resulting manufactured goods to be moved out to consumers. In the period following the World War II, there arose a *service* economy, which allowed production and marketing to become progressively more personalized. The service economy is currently focused on devising ways to support and to disseminate services electronically, including services delivered not only to consumers and producers but also to all parts of society, including medicine.

We have now reached the point at which cloud computing and other types of advanced infrastructure are bringing about a situation in which knowledge objects can be delivered in an efficient manner to those who need to consume them. And just as highways were the infrastructure necessary for a manufacturing economy, serving as the arteries along which raw materials and manufactured goods coming in from all directions could flow, so we believe that ontologies will in the future provide an important pillar in the infrastructure that is necessary for the accurate and safe delivery of all varieties of knowledge.

Examples include the way in which predictive analytics based on machine learning are now being increasingly approved for use by the Food and Drug Administration (FDA). If our hypothesis is correct, then this and many other types of innovative software will be increasingly common at all stages and levels of hospital care, built

P. L. Elkin (✉)
Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA
e-mail: elkinp@buffalo.edu

B. Smith
Department of Philosophy, University at Buffalo, Buffalo, NY, USA

Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA
e-mail: phismith@buffalo.edu

P. L. Elkin (ed.), *Terminology, Ontology and their Implementations*, Health Informatics, https://doi.org/10.1007/978-3-031-11039-9_24

on the basis of ontology-tagged data. We now have data showing that machine learning algorithms built off of ontology terms are more accurate and more stable in the face of missing or bad data in the training set [1].

We believe that the reach and accuracy of genuinely useful machine learning algorithms can be combined with deterministic models involving the use of ontologies to enhance these algorithms with prior knowledge. The ontologies themselves—after the model of the Open Biological and Biomedical Ontologies Foundry [2]—will be built by experts in the corresponding domains and maintained in such a way as to keep pace with the advance of science while at the same time preserving consistency and adapting to each new advance in the software and hardware that drives information.

This foundation in ontologies will improve evaluation of the algorithms, for example by improving discoverability of relevant data pertaining to the outcomes generated through their use.

The foundation in ontologies will also enable the implementation of algorithms in different disciplines and communities to become more and more consistent over time—based not least on the fact that consistency of ontologies (and to a degree of ontology-tagged data) can be tested at the click of a mouse. And as all terms in the ontologies have formal definitions, everything that happens to the patient becomes more tightly and consistently captured—and more easily analyzable—than is possible where we are dealing, for example, with text in clinical notes.

The evaluation of algorithms, like all other forms of evaluation in the medical enterprise, is a continuous process. The algorithms themselves cover a domain which stretches all the way from genes and proteins to phase IV clinical trials. By being rooted at each level and at each stage in a framework of carefully curated and continuously updated ontologies, the entire process of algorithm and ontology development will improve in both reach and consistency over time. Algorithms and ontologies will be improved in lockstep, leading to an improvement of care and enhancement of medical knowledge at all levels.

We are herewith imagining a world in which clinicians have at their fingertips the most accurate possible knowledge at every stage of patient care. The knowledge economy that will allow this vision to become reality will require strong governance of both data and the personnel who supply and use them. Hospital management, too, will have at their fingertips the most accurate possible knowledge that they need to run the hospital. All of this will require agreement on the meanings of the terms that are used at all levels, and these definitions, too, are to be provided by ontologies.

We require data governance to define the processes that we use to build our algorithms and the methods we use to apply them. Multiple different sorts of actors will be involved in this process, and once again the ontologies will allow us to define not only the actors but also their respective roles and responsibilities and how they are executing these roles and responsibilities at any given stage, not least in the maintenance and application and testing of ontologies.

There are some who embrace a vision of using software to create what are called digital twins—of aircraft, of cities, and of rivers—and also hospitals [3]. The idea is that all the data captured about the hospital and its processes will be used to create a digital simulacrum of the hospital itself that models all the events in the hospital from minute to minute. Our vision is, certainly, along these lines. But it is more sophisticated, since it takes account from the start of the fact that there will be multiple interacting knowledge layers in the hospital, just as there are multiple interacting layers of responsibility and control. Adding new software into a system such as this results in manifold changes in the system, and this applies too of software under the heading of "digital twin"—the hospital digital twin will *change* the hospital itself and will therefore necessarily and from the very start fall short of its goal of being a digital twin.

We, in contrast, recognize the dependencies between different data flows and data operations as we track both doctors and patients through the successive processes of care and treatment. We can define dependencies, and we can track through referents (uniquely identified instances of types [4]) the connections between facts. We can link levels of certainty to these facts, and we can then operate formally on these definitions to ensure ontologically consistent representations which take account of uncertainty and other error factors. Ontologies will be applied also to the consistent representation of data provenance, so that we can track all changes to data as input to knowledge objects or predictive analytic training sets. We see this as a way to register objects for reuse. Through this reuse, we once again promote consistency between researchers and developers, following the principle that the more eyes on a given piece of data or on given treatments of data in a hospital system, the more rapidly will errors be detected and systems correspondingly corrected.

A world with fewer errors and an improved understanding of data and knowledge products will lead to better and more efficient lives for everyone, and not only those who become entangled in hospital processes. An example of how this might work in the field of predictive analytics might run as follows: An approved algorithm can be annotated with metadata describing the clinical and molecular properties of the patients in the training data. If these metadata are encoded using ontologies, a clinician can query the metadata to determine whether there is a match with a patient's profile in order to determine whether the algorithm can be used on the patient he/she is currently treating (therein keeping patients safe while applying useful knowledge-based tools and methods).

What Does the Government Need to Do to Support Our Knowledge-Based Economy?

Governments need to support the development and maintenance of public domain ontologies in all medically relevant fields, and indeed in all fields that can support our knowledge-based infrastructure needs. The high-quality interoperable ontologies we envisage are important in ways comparable to the Interstate Highway System, and then the potential benefits to be gained from supporting such development are clear.

It is clear also why we cannot easily perform the corresponding cost-benefit calculations, just as there is no way in which we could calculate the benefits gained from the building of the Interstate Highway System by somehow comparing the world with the system as it is today with the world as it would have been without it. Currently, where medical terminologies and ontologies are too often locked in proprietary siloes and have other features which degrade interoperability, public support for an all-encompassing, public domain ontology infrastructure will unleash the power of *all* the data we have at our disposal through tagging with ontologies. The data themselves, insofar as they relate to individual patients, will remain behind HIPAA firewalls. But with the proper protections, potentially involving application of blockchain technology, they can still be managed as if they were part of a single data environment.

The results will empower business and service providers to be successful in their competition within this knowledge-based economy. Leaders that embrace these ideals and create the infrastructure will see their economies flourish and those that do not will compete poorly in the future.

In the United States, the National Library of Medicine has done a great job in cataloging and delivering clinical terminologies. The Department of Veterans Affairs has done increasingly strong work in the development and implementation of clinical ontologies, through the SOLOR program [5]. More recently, the Department of Defense (DoD) and the Intelligence Community in the United States have begun to adopt and to disseminate ontology models and principles, and the International Standards Organization and the International Electrotechnical Commission have established a series of ontology standards under ISO/IEC: 21838.

What Would We Like to See in the Future?

We would like to see a national dissemination team for ontology. We would like to see the resultant ontologies authored by the best experts in the world, for example in exchange for academic credit toward promotion and tenure. We would like to see implementation teams that devise strategies that make it easier to build systems and knowledge objects with rather than without ontology. We anticipate that ontology-based systems will have a lower total cost of ownership than those that are not ontology based. This is because the highest costs of implementation and maintenance are often due to system inconsistencies, which appear only after the system has been brought into use. Rectifying the resultant problems will require considerable effort, often involving additional training.

Ontology authorship requires ontologists, informaticians, and subject matter experts to design, author, test, and maintain ontologies and to use ontologies in tagging data and building information architectures, which can successfully use the tagged data and feed the results back into the ontology development process. This requires stable long-term funding to encourage people to choose an operational career in ontology. One way to populate this workforce is for example from schools of library science as well as from programs in information and computer science and indeed philosophy.

Currently, many ontologists have been trained to view the results of their work as responses to customer requirements. It will be clear that, from the perspective advanced here, this practice should be abandoned, since it yields ontological redundancy. Rather, ontologists should be trained as scientists are trained, which means that we need ontology (science [6]) based on commonly accepted principles and best practices, and a cumulating body of commonly accepted scientific knowledge. A standardized repository of high quality ontologies for the life sciences can be found uniquely in the OBO foundry [7]. But more, many more, ontologies will be needed in the future, including ontologies addressing new kinds of challenges [8].

Conclusions

Ontology is a mature discipline that will prove essential for our knowledge-based economy. Federal sources should consider this infrastructure as an essential component of a healthy economy. We need further development of our ontology workforce and sustainable strategies to advance best practices in this field in order to continue leading the world in knowledge-based products and services.

References

1. Mullin S, Zola J, Lee R, Hu J, MacKenzie B, Brickman A, Anaya G, Sinha S, Li A, Elkin PL. Longitudinal K-means approaches to clustering and analyzing EHR opioid use trajectories for clinical subtypes. J Biomed Inform. 2021;122:103889. https://doi.org/10.1016/j.jbi.2021.103889. Epub ahead of print. . PMID: 34411708
2. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007;25(11):1251–5.
3. Karakra A, Fontanili F, Lamine E, et al. Pervasive computing integrated discrete event simulation for a hospital digital twin. In: IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA); 2018. p. 1–6.
4. Ceusters W, Smith B. Strategies for referent tracking in electronic health records. J Biomed Inform. 2006;39(3):362–78.
5. Seftel-Kirk H. VA's Solor – bringing seamless care to veterans and beyond. FedHealthIT. 2019. https://www.fedhealthit.com/2019/04/vas-solor-bringing-seamless-care-to-veterans-and-beyond/.
6. Smith B. Ontology (Science). Formal Ontology in information systems. Amsterdam: IOS Press; 2008. p. 21–35.
7. Jackson RC, Matentzoglu N, Overton JA, et al. OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. bioRxiv. 2021.
8. Limbaugh D, Landgrebe J, Kasmier D, et al. Ontology and cognitive outcomes. J Knowl Struct Syst. 2020;1(1):3–22.