

Does not compute: challenges and solutions in managing computable biomedical knowledge

David Wong,^{1,2} Niels Peek^{1,3}

To cite: Wong D, Peek N. Does not compute: challenges and solutions in managing computable biomedical knowledge. *BMJ Health Care Inform* 2020;**27**:e100123. doi:10.1136/bmjhci-2019-100123

Received 08 December 2019
Revised 25 February 2020
Accepted 20 March 2020

ABSTRACT

Computers can potentially play a key role in resolving knowledge mobilisation bottlenecks in health and care through decision support at the point of care based on computable biomedical knowledge (CBK). But the management of CBK comes with a range of significant computer science challenges. Some of these have been suitably addressed through the development of CBK methods and tools, while others require further research and development. We review the main challenges associated with creating, reasoning with and sharing CBK, and describe current state-of-the-art solutions as well as outstanding issues. We argue that a radical approach, in which all evidence generation is suitable for computation at the outset, is ultimately needed to take full advantage of CBK.

INTRODUCTION

Conventionally, knowledge is expressed in words, symbols and pictures and disseminated through books, journals and papers. Interpretation, manipulation (such as summarising) and acting on such knowledge require that a person reads that book, journal or paper—a slow and laborious process. This is the main bottleneck for mobilisation of the rapidly growing volume of biomedical knowledge.¹ Computable knowledge is knowledge expressed as computer code: machine-interpretable statements that are inaccessible to direct human comprehension. Since computers can interpret, manipulate and reason with computable knowledge, this can potentially partially resolve the knowledge mobilisation bottleneck.

The management of computable knowledge comes with a range of computer science challenges—some of which have been suitably addressed through the development of methods and tools, while others require further development. The purpose of this short report is to provide an overview of the computer science challenges in creating, managing and mobilising computable biomedical knowledge (CBK).

CREATING COMPUTABLE KNOWLEDGE

Computable knowledge is created through the development of computer-interpretable objects (and relationships between them) in a way that fully and unambiguously captures the knowledge in a given source. For instance, we may take the National Institute for Health and Care Excellence (NICE) guidance for recognising and responding to deterioration in acutely ill adults in hospital (<https://www.nice.org.uk/Guidance/CG50>) and convert that into a fully computer-interpretable algorithm that can be subsequently used as the basis for clinical decision support.

The process of converting existing knowledge into computable form is called ‘knowledge formalisation’, and it is never straightforward for three reasons. First, it requires that we make explicit all assumed background knowledge and ‘common sense’ that people draw on when interpreting the source knowledge. Second, all forms of ambiguity should be removed to enable machine interpretation. Third, for verification and maintenance purposes we need a clear correspondence between the knowledge source and its computable form.

Ambiguity arises sometimes due to limitations of natural language, but it may also be due to oversight or assumed knowledge. Clinical guidelines are typically written for an audience in which baseline clinical knowledge can be assumed. Otherwise, ambiguity may be introduced intentionally to enforce generalisability. For example, the aforementioned NICE guidance states that ‘in specific clinical circumstances, additional monitoring should be considered’ but deliberately does not define or provide examples of relevant circumstances. In any case, the presence of ambiguity means that knowledge cannot be easily translated into a computable form.

Computer scientists have developed bespoke computer languages and tools to create CBK. Examples are the Arden syntax



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Centre for Health Informatics, School of Health Sciences, The University of Manchester, Manchester, UK

²School of Computer Science, The University of Manchester, Manchester, UK

³NIHR Manchester Biomedical Research Centre, The University of Manchester, Manchester, UK

Correspondence to

Prof Niels Peek;
niels.peek@manchester.ac.uk

for representing event-condition-action rules (a specific type of IF–THEN rules)² and Protégé,³ the most widely used software for building and maintaining ontologies. There also exist intermediate knowledge representation that facilitates the process of converting natural language guidelines into computable form.⁴

In the process of making background knowledge explicit and resolving ambiguities, the relationship between knowledge source and computable object can become blurred. For instance, one might decide to include, in the computable guideline, clear-cut criteria for the circumstances in which additional monitoring should be considered based on the consultation of experienced critical care clinicians—something that was not specified in the source guideline. This would improve the ability to provide actionable decision support but reduce the correspondence between the computable object and source guideline. One possible solution is to develop the source guideline and computable guideline concurrently. This approach has been trialled with some success in the Netherlands where Goud *et al*⁵ developed a computerised clinical decision support system alongside the development of a new version of national clinical practice guidelines for cardiac rehabilitation.

INFERENCE

Once biomedical knowledge is available in computable form, computers can mobilise that knowledge. For instance, it enables more precise searches for relevant knowledge than is currently possible through clinical databases, because queries would no longer depend on imprecise natural language terms. But the most powerful way to mobilise knowledge is through point-of-care computerised decision support. This requires the manipulation of computable knowledge in a meaningful way to produce actionable outputs, a process called ‘inference’.

Inference with CBK requires some form of logical or probabilistic reasoning in which persistent knowledge such as computerised clinical guidelines is combined with specific data from individual patients. For instance, we might want to assess whether an individual patient admitted to hospital requires additional monitoring. This would involve assessing each patient’s record against the criteria for additional monitoring. There exist many software packages for this type of inference with IF–THEN rules (eg, Karadimas *et al*⁶) and with formal ontologies.⁷

Things become more challenging when multiple knowledge sources are relevant for a given case. For instance, often multiple guidelines will be applicable for a patient with multimorbidity. This requires meta knowledge to resolve conflicts, something which is far from trivial. Further challenges arise when considering the veracity of knowledge. For instance, we may consider current clinical guidelines produced by NICE to be more trustworthy than information from social media. For computers to make such assessments, we require

specification of meta-data such as the date of publication and the organisation that produced the knowledge. We also require accompanying meta-knowledge that allows the computer to involve meta-data in inference processes. When there are no guidelines, it may be needed to automatically synthesise clinical research that requires methods to interpret and understand the results of published clinical studies. There are emerging methods and tools for all of this, but at this moment in time none of them is mature enough for routine deployment.

SHARING COMPUTABLE KNOWLEDGE

Early languages for describing CBK objects struggled with dependencies on local terminology and data sources. In the Arden syntax, this was known as the ‘curly braces’ problem²: expressions referring to data sources would be written between curly braces but they would be completely dependent on the local database schema. This meant that knowledge objects described in the Arden syntax could not be shared between providers. Every provider had to go through their own knowledge formalisation process. The old knowledge mobilisation bottleneck had been replaced with a new one.

Since the early 2000s many efforts have focused on creating standards for sharable computable knowledge, such as the Guideline Interchange Format.⁸ Mobilising CBK on a large scale requires standard approaches to representing clinical knowledge in both human-readable and machine-executable formats, as well as standard approaches for leveraging CBK to provide decision support across different applications and care settings.⁹ In recent years, steps have been made towards interoperable integration of decision support with electronic health records using HL7 Substitutable Medical Applications, Reusable Technologies on Fast Healthcare Interoperability Resources (SMART on FHIR).¹⁰

CONCLUSION

Important progress has been made in developing methods and tools for creating, managing and mobilising CBK, but significant challenges still exist. These challenges might be surmountable only through radical formalisation of the biomedical knowledge management process.¹¹ In part, evidence-based medicine research has made steps towards this by standardising how evidence is reported and synthesised via organisations such as EQUATOR Network and the Cochrane Collaboration. But arguably we can only expect to resolve the significant biomedical knowledge mobilisation bottlenecks that still exist when computable objects are generated, transferred and interpreted at each stage of the knowledge management process. In this approach, all evidence generation would be suitable for computation from the outset. The natural language text describing the experiment and the outcome (ie, the academic paper) would

be a surface human-readable representation. The paper would be supported with a set of results in a computable format that could be further processed to yield higher level information such as systematic reviews and clinical practice guidance—all available on demand through fully automated inference.

Twitter David Wong @drdavecwong

Contributors The manuscript was drafted by DW and NP and subsequently reviewed by both authors.

Funding This research was partially funded by the National Institute for Health Research (NIHR) Manchester Biomedical Research Centre (IS-BRC-1215-20007). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- 1 Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7:e1000326.
- 2 Samwald M, Fehre K, de Bruin J, et al. The Arden SYNTAX standard for clinical decision support: experiences and directions. *J Biomed Inform* 2012;45:711–8.
- 3 Musen MA, Protégé Team. The Protégé project: a look back and a look forward. *AI Matters* 2015;1:4–12.
- 4 Hajizadeh N, Kashyap N, Michel G, et al. Gem at 10: a decade's experience with the Guideline elements model. *AMIA Annu Symp Proc* 2011;2011:520–8.
- 5 Goud R, Hasman A, Strijbis A-M, et al. A parallel guideline development and formalization strategy to improve the quality of clinical practice guidelines. *Int J Med Inform* 2009;78:513–20.
- 6 Karadimas HC, Chailloleau C, Hemery F, et al. Arden/J: an architecture for MLM execution on the Java platform. *J Am Med Inform Assoc* 2002;9:359–68.
- 7 Tsarkov D, Horrocks I. *FaCT++ description logic Reasoner: system description*. Berlin, Heidelberg: Springer, 2006: 292–7.
- 8 Boxwala AA, Peleg M, Tu S, et al. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *J Biomed Inform* 2004;37:147–61.
- 9 Kawamoto K, Del Fiol G, Lobach DF, et al. Standards for scalable clinical decision support: need, current and emerging standards, gaps, and proposal for progress. *Open Med Inform J* 2010;4:235–44.
- 10 Mandel JC, Kreda DA, Mandl KD, et al. Smart on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc* 2016;23:899–908.
- 11 Bechhofer S, Ainsworth J, Bhagat J. *Why linked data is not enough for scientists*, in: *2010 IEEE sixth Int. Conf. e-Science*. IEEE, 2010: 300–7.