

Briefing paper on: Knowledge Infrastructure Requirements for Computable Biomedical Knowledge (CBK)

Briefing Advisors: Bob Greenes and Carl Lagoze

Briefing Paper Authors: Bob Greenes, Carl Lagoze, Beltran Figueroa, and Allen Flynn

What is meant here by the term “knowledge infrastructure”?

DEFINING KNOWLEDGE INFRASTRUCTURE

Here we suggest that a *general platform* of some sort is needed for creating, representing, managing, stewarding, applying, and using knowledge in the form of CBK. We claim the need for such a general platform for CBK, even given the wide range and variation of kinds of knowledge that exist and can be made computable. Although the term “knowledge infrastructure” is not easily defined,^{1,2} we define it, quite loosely, in terms of a platform to make it easier to perform the tasks listed above. This paper is intended to stimulate sharing of a variety of perspectives about the meaning of the term “knowledge infrastructure”, as a platform that supports CBK by providing specific capabilities.

NATURE OF CBK

The scope of this paper is limited to knowledge in the form of computable biomedical knowledge (CBK). CBK is computable in the sense that it is *machine-executable*. Cutting to the chase, CBK is biomedical knowledge, broadly construed, that is represented in a form allowing it to be used either *to instruct* a computing machine directly or *to structure instructions* that are then used by a computing machine. CBK can be of a variety of types, such as executable production rules, semantic reasoning, formalized queries, or application of analytics such as predictions. In the context of CBK, it is reasonable to question the distinctions between, and definitions of, data, information, and knowledge. However, here we sidestep this by claiming that any content that can be formally described and applied computationally to support inference or decision-making, or to support cognition, can be considered knowledge.

KNOWLEDGE INFRASTRUCTURE SUPPORTING CBK

As the term is used here, “knowledge infrastructure” indicates any underlying, general-purpose, domain-agnostic, scalable capability to decrease friction that impedes CBK sharing.² Knowledge infrastructure is sociotechnical.¹⁻³ This brief paper focuses on knowledge infrastructure requirements either for technical systems, such as digital libraries, knowledge-bases, rules engines, search engines, publishing systems, and IT service buses, or for implementations of architectural patterns that are service-oriented, event-driven, agent-based, etc.⁴

Thought Questions:

Do the definitions of CBK and knowledge infrastructure above ‘fit’ with your understanding of these things?

Why or why not?

INTRODUCTION: Envisioning a CBK Pipeline as Infrastructure for the Learning Health System

To organize this paper, we invoke the notion of a “CBK pipeline” as infrastructure. As illustrated in Figure 1, the CBK pipeline we have in mind can be overlaid on a model of a learning cycle for the Learning Health System (LHS).³

Knowledge infrastructure must support the CBK pipeline. The pipeline stretches from near the end of the “Data to Knowledge” section of the learning cycle (blue) to near the end of the “Knowledge to Performance” section (red).

The CBK Pipeline has three high-level sequential stages of work, each involving a host of complex, iterative work processes and related infrastructure requirements. These stages are: (1) **Create &**

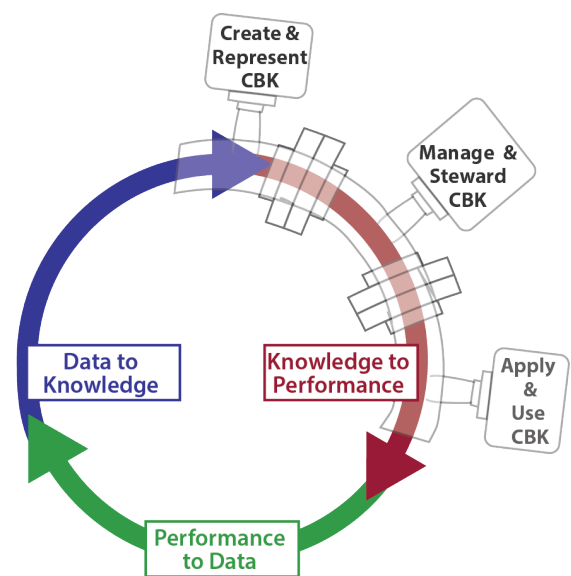


Figure 1. Three-Part CBK Pipeline Overlaid on a Learning Cycle

Represent CBK, (2) **Manage & Steward CBK**, which includes curation, updating, and making it FAIR (as detailed below), and (3) **Apply and Use CBK**, which includes technical deployment and engagement of CBK by end users to get advice. This partial pipeline does NOT encompass the full learning cycle, but when connected with other components, it could. Also, learning cycles apply to improving the processes that take place at *any point* on this cycle, reflecting the “fractal” nature of this model of ongoing learning.

STAGE 1: Requirements to Provide Needs-based Support for CBK Creation and Representation

The **Create & Represent** CBK stage of the CBK pipeline (Figure 1) involves tools, processes, and workflows needed after a community has completed empirical *analytic and/or deliberative* “data to knowledge” work, resulting in new knowledge. In other words, the infrastructure requirements that pertain at this stage are *post-discovery* requirements. If the community desires to share its new knowledge, then it will need some means for doing so, with increasing capability depending on how widespread the sharing is intended to be. Growing interest in CBK sharing by communities of health stakeholders motivates support for CBK **creation** and **representation** to meet their needs.

Here, “to create” refers to the act of using a process that directly results in CBK, whether manually or automatically. This could be as simple as saving a file as CBK within a statistical software program. In contrast, “to represent” means to convert knowledge-in-the-world into CBK via knowledge engineering.

Similar requirements for technical knowledge infrastructure apply to creating and representing CBK. These begin with the need for some technical specifications for CBK. In addition to these specifications, reliable methods and tools for (a) creating CBK automatically, (b) readily converting bodies of knowledge-in-the-world to consistently represented CBK, and (c) engineering CBK piece-wise are all required. One can imagine software and hardware that supports rapid CBK production along these lines.

Once CBK is created, the need to update it becomes immediately apparent, and is further elaborated in the next stage. The ongoing flow of new biomedical knowledge,^{1,3} i.e., the results from the data to knowledge process in Figure 1, creates the need to constantly expand and refine CBK.⁵ Hence, adding to the requirements for CBK knowledge infrastructure are tools for editing and versioning CBK.

In addition, there are key aspects of publishing and certification that directly link to CBK creation and representation. One thing that is needed is a, “tool that facilitates knowledge editing and exchange between knowledge creators/publishers and integrators.”⁵ Just as important will be components of infrastructure to support the tasks of testing, verifying, and validating CBK. These could include large synthetic data sets for testing, secure and reliable certification mechanisms, CBK performance reporting systems, and related support for reviewing CBK.

STAGE 2: Requirements to Support CBK Management and Stewardship

When applied to developing a general-purpose knowledge infrastructure for CBK, the “FAIR” principles serve as a high-level description of what the infrastructure must bring about, which is to make CBK *Findable, Accessible, Interoperable, and Reusable* at scale. The tasks of “**management and stewardship**” have been previously related to the “FAIR” principles, as they apply to data and data sets.⁶

For CBK to be findable, infrastructure requirements must support its ready discovery. Discovery at scale is non-trivial and depends on rich metadata being applied to instances of CBK.² These metadata must describe features of interest to a diverse audience of researchers, librarians, informaticians, developers, clinicians, and consumers.^{3,4} In addition, discovery requires keeping updated records on CBK branching, which is when a CBK instance is further developed in more than one way, CBK splitting, which is when a CBK instance is subdivided into multiple sub-instances, and on CBK citations too.

For CBK to be accessible once it is found, mechanisms to support authentication and authorization to CBK archives come into play, as does tracking of the whereabouts of CBK over time. For a robust, “deep”⁶ infrastructure, CBK stewardship requirements extend to incorporate best practices for digital preservation, which include a suite of tools, operations, standards and policies.⁷

In support of reusability, once CBK is available, licenses, utilization details, and related safety and impact metrics will need to be collected and shared.⁵ In addition, support for curating CBK collections, some of which will help compose or combine instances of CBK for enhanced use, will be needed.³

STAGE 3: Requirements to Apply and Use CBK in Real-World Workflows within Health Applications

The need to “**apply and use**” CBK, under local conditions, leads to infrastructure requirements to support its contextualization, customization, and interoperation.⁵ In this case, to apply (or deploy) and use CBK is to combine it with *instance data* to generate a message of advice, or guidance, which is then transmitted to a person. One capability of great interest is to apply and use CBK in context, similar to the way GPS-enabled smart navigation provides information about their surroundings to people in motion. This goal of having CBK-enabled systems tell people about their health, and that of others, while they are having clinical, wellness, and other life experiences, has been articulated.⁵ Knowledge infrastructure requirements for in-context health advisories demand “plug-n-play” interoperability for CBK.

Knowledge infrastructure for CBK must support CBK use in many other ways besides in-context use. For example, to go from “Performance to Data” (Figure 1) requires gathering relevant data about CBK use, along with the impacts and outcomes that follow from its use, so the CBK can be evaluated and improved. Also, as updates to CBK occur, a platform must provide mechanisms to support tracking and notification so that users of CBK are systematically made aware of changes.

ADDENDA: Anticipated Challenges to Further Advancing Technical Knowledge Infrastructure for CBK

Out of many, two anticipated challenges for effective CBK knowledge infrastructure are noted here.

First is the challenge of having too many competing efforts thwarting the ultimate convergence that is needed to arrive at a general-purpose knowledge infrastructure. A balance must be struck by a CBK community between common, workable solutions that deliver on stable core requirements and “innovation at the edges”⁸, which is needed to further advance knowledge infrastructure for CBK.

Second there is the challenge of securing CBK so that it is not corrupted, intentionally or unintentionally. A variety of technical methods and policies may help address the CBK security challenge.

CONCLUSION

Knowledge infrastructure requirements for a general-purpose, domain-agnostic infrastructure for CBK span a system that supports creation, representation, management, stewardship, application (or deployment), and use of CBK. The over-arching goal of such an infrastructure is to make it much easier to do all of the work in these areas so that rapid, widespread sharing of CBK results. When this happens, learning cycles within the LHS can be greatly facilitated.

REFERENCES

1. Edwards, P.N., et al., 2013. Knowledge infrastructures: Intellectual frameworks and research challenges.
2. Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C. and Borgman, C.L., 2011. Science friction: Data, metadata, and collaboration. *Social Studies of Science*, 41(5), pp.667-690.
3. Friedman, C.P., Rubin, J.C. and Sullivan, K.J., 2017. Toward an Information Infrastructure for Global Health Improvement. *Yearbook of Medical Informatics*, 26(01), pp.16-23.
4. Greenes, R. “Clinical Decision Support 2nd Edition: The Road to Broad Adoption”. CHAPTER 28 “A Clinical Knowledge Management Program”. Rocha, R.A. Maviglia, S.M., Sordo, M., Beatriz H. Rocha, B.H.
5. Greenes, R.A. Peleg, M. Rector, A., Osheroff, J. “Reusable Knowledge for Best Clinical Practices: Why We Have Difficulty Sharing And What We Can Do”.
6. Wilkinson, M.D., et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, p.160018.
7. Conway, P., 2010. Preservation in the age of Google: Digitization, digital preservation, and dilemmas. *The Library Quarterly*, 80(1), pp.61-79.
8. Sidak, J.G., 2006. A consumer-welfare approach to network neutrality regulation of the Internet. *Journal of Competition Law & Economics*, 2(3), pp.349-474.