



Briefing paper on: Metadata Standards for Computable Biomedical Knowledge (CBK)

Briefing Advisor: Rachel Richesson

Briefing Paper Authors: Brad Iott and Allen Flynn

How one instance of CBK could have many uses in a Learning Health System

CREATING AN INSTANCE OF CBK

Remy is a data scientist. She creates predictive models using powerful analytic techniques. Recently, using a large patient data set with many biomarkers, Remy created a highly accurate model to predict recovery times after hip replacement surgery. This model is an instance of CBK.

PUBLISHING CBK

Remy worked with Kyle to publish her model's "R code" with a detailed description. Together, following a step-wise process, they tagged her "R code" with an assortment of metadata, verified it was virus-free, certified that it worked as designed (using synthetic data), and stored it in an archive.

CBK USE IN CLINICAL PRACTICE

Carol is an orthopedic surgeon. She performs many hip replacements. Carol did an online CBK search to find models to predict hip replacement surgery recovery time. She found Remy's model. She worked with her organization to implement Remy's model in their EHR. Now Carol uses patient-specific recovery time predictions, calculated by the EHR, during her pre-surgical consultations with patients.

CBK USE FOR POPULATION HEALTH

William is a biostatistician. He works for a large insurance company. William also searched for CBK and found Remy's predictive model. William deployed Remy's model in his population health analytics suite and used it to upgrade an *expected cost of care analysis* for insureds with hip problems.

CBK USE IN BIOMEDICAL RESEARCH

Finally, Darla is a bioscientist who studies the chemistry of muscle tissue regeneration. Darla wanted tissue samples from patients exhibiting slow muscle regeneration after surgery. She too found Remy's predictive model and contacted Remy to discuss it. Darla used Remy's model to help her identify candidate study subjects for tissue collection. Now Darla and Remy are collaborators on new research to discover the mechanisms that govern the pace of muscle regeneration after hip surgery.

Thought Questions:

1. What metadata are needed for Carol, William, and Darla to find, access, deploy, and reuse Remy's CBK?
2. When Remy's model, an instance of CBK, is improved, how might these stakeholders be informed?

THEME 1: Importance of Establishing a Reliable, Ongoing Process for Developing Metadata Standards

CBK is critical to the Learning Health System (LHS).(1) In a sense, *archives of CBK* must serve a "memory function" for the LHS, enabling the LHS to know what it has learned at any point in time.

For CBK to function effectively in this way, Metadata standards[◇] are essential (2). Metadata are descriptive data that support CBK FAIRness, meaning that they help ensure CBK is *findable, accessible, interoperable, and reusable* (3). However, developing metadata standards that actually get used is difficult and costly (4). In part this is because standards are never finished (2). Standards either evolve or wither (2). A thoughtful approach to ongoing CBK metadata standards development, including painstaking prioritization of effort, is required.(2) A reliable, ongoing process of CBK standards identification and development is needed.(2) To this point, as much as possible, it is wise to find and adopt existing standards for CBK metadata (5). Nevertheless, in some instances, new metadata standards for CBK may be needed.

[◇] By metadata standards, we mean *data standards of different sorts* that are used to *describe instances of CBK*.

Assuming some new CBK metadata standards are needed, a reliable, transparent standards development process should seek to achieve *rough, broad-based consensus* and not get bogged down in minutiae (4). Initiation of an ongoing process of developing new CBK metadata standards involves identifying a wide range of stakeholders for whom a new standard is relevant (4). Once a community forms, there is much work involved to achieve community consensus. A starting point for this work can be the determination of common data elements for any new CBK metadata standard (5).

Thought Questions:

1. What are examples of existing standards that can be adopted for CBK metadata, such as Dublin Core or FHIR?
2. What may be new CBK metadata standards needing to be developed “from scratch”?
3. In general, who should be included in a list of stakeholders for CBK metadata standards? Why?

THEME 2: Mechanisms to Build and Achieve Rough, Broad-Based Community Consensus

Achieving rough, broad-based community consensus is essential for the development of new CBK metadata standards (2). For CBK, as for the LHS, the notion of community is expansive and inclusive of many and diverse stakeholders. Various stakeholders and users are likely to be interested in different facets of CBK. Their collective input on CBK metadata standards is important (2).

As the vignette suggests, there will be different CBK stakeholder communities. It is necessary to account for the needs of researchers who generate CBK, librarians who curate it, developers who deploy it, and end-users who benefit from applying it to their work. Folks in these and other roles are certain to have role-oriented perspectives on what it takes to make CBK “FAIR” *for them*.

There is the potential to adopt the processes of other successful open, consensus-driven standards development efforts for development of CBK metadata standards. Those involved in the process must first determine the scope of the standard (4). Next, a draft of the standard may be written. The draft can then be made available for comments and balloting through a consensus-based process run by an accredited standards development organization (SDO). In consensus-based SDOs, issues raised in negative ballots must be addressed before final voting on the standard takes place. Once the draft is finalized, it may be implemented for testing. Issues that arise during testing can then be corrected.

ANSI accredits SDOs and offers guidance on due process for standards development (6). Due process promotes consensus of the full community while avoiding dominance by vocal or well-funded members. Community consensus is facilitated by having written operating procedures for harmonization with existing standards, public notification, open voting, and handling of appeals (6).

Thought Questions:

1. What does the idea of “*rough, broad-based community consensus*” mean to you?
2. Is a role-oriented approach to identifying CBK metadata needs and scoping CBK metadata standards workable and useful? Why or why not?

THEME 3: Determination of Common Data Elements as a Starting Point for CBK Metadata Standards

Data elements are precisely defined units of data with known attributes (ISO 11179-3). An example of a data element is ‘title’, defined by Dublin Core as a “name given to a resource” formatted as an `rdfs:literal` string (7).

Common data elements (CDEs) are those that can be uniformly represented to bring value across many domains. To help make resources “FAIR”, the scope of CBK metadata standards may start to be realized by defining common data elements (CDEs) or classes of CDEs.

To avoid re-work and limit confusion, CDEs for CBK can be drawn from existing standards through a process of incorporating and harmonizing what exists with what is needed. The degree to which new CDEs are needed for CBK has to be determined following exploration of existing metadata standards. In addition, it is possible to use data-driven approaches to determine what are wanted CDEs to make CBK “FAIR.” For example, systematic analyses of the CDEs used today in bibliometric, scientometric, knowledge analytics, and other related scientific domains may reveal high-value CDEs for CBK (2).

Once CDEs are determined and documented, they may be vetted by communities, via expert review and validation, before they are incorporated into a new standard. This approach leverages the intersecting expertise of various stakeholders invested in each CDE to help ensure the consensus is achieved. Given the multidisciplinary nature of CBK, an obvious goal is to bring about CBK metadata standards that have relevance for a wide range of users and use cases.

THEME 4: Anticipated Challenges to Advance CBK Metadata Standards

Besides some challenges that have already been mentioned, other challenges are anticipated in the pursuit of CBK metadata standards.

Scope creep is a significant concern. It is the natural tendency for the scope of a metadata standard to expand beyond what was initially planned. A CBK community will need to pursue work towards standards that help make CBK “FAIR” methodically to guard against this. Standards created before minimally sufficient sets of requirements can be determined are especially likely to result in scope creep. An ongoing challenge is to do standards development work in an organized, agile and iterative way to limit scope creep while evolving a standard effectively.

The work of implementing standards in the real-world is substantial, and sometimes too much to overcome (2). As CBK itself evolves over time, the metadata standards for CBK will have to evolve. This adds to the implementation burden and points to the need for processes for standards implementation.

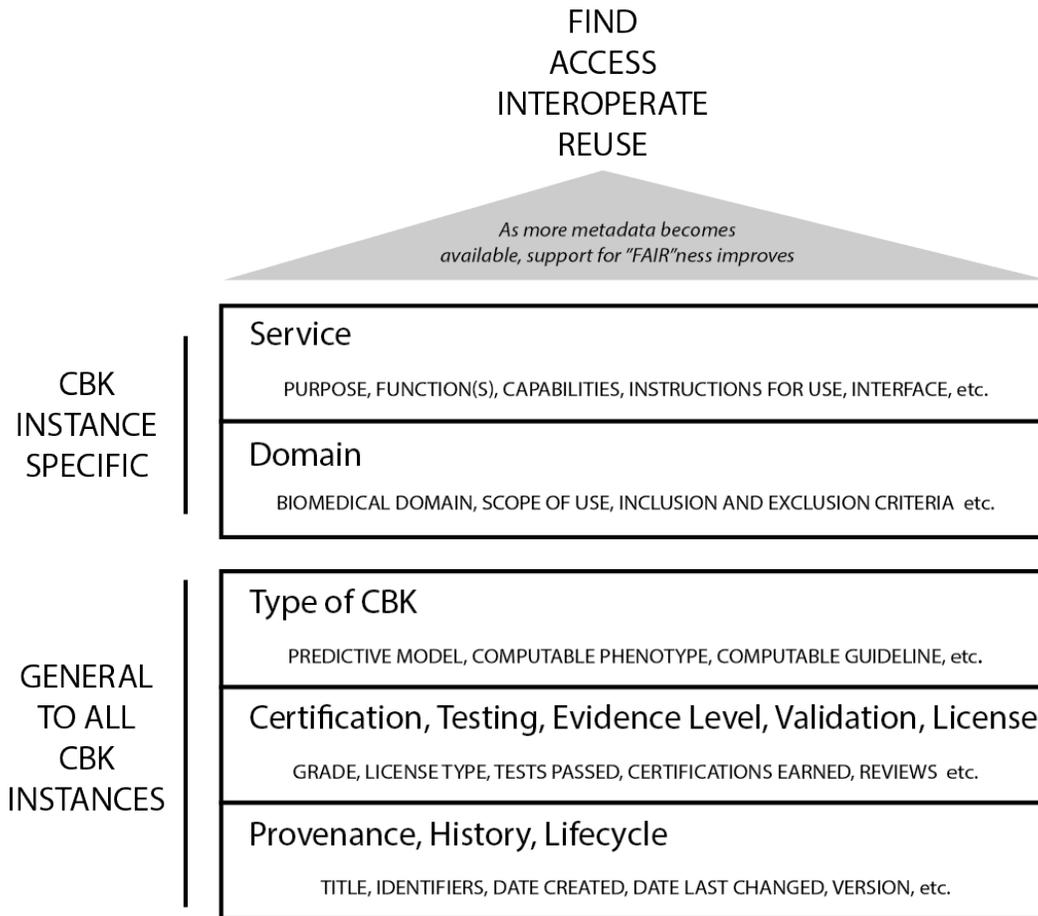
Finally, the creation of a culture within a community focused on robust metadata standards to make CBK “FAIR”, within the context of the Learning Health System, will require a paradigm shift away from standards development towards evolutionary processes for standards maintenance and improvement. This shift will need to be supported by infrastructure that continuously enables CBK-related best practices to be discovered, applied, evaluated and improved to make CBK fit for use.

CONCLUSION

Within the larger context of creating the LHS, a reliable, ongoing process of CBK metadata standards development, with ongoing maintenance and improvement, is needed. The purpose of this is to make CBK “FAIR” for a wide and diverse audience of users. To be effective, these ongoing processes will have to organize a broad, diverse community, spanning many roles, to achieve rough consensus. The community that assembles for this purpose will need to prioritize its efforts carefully. After identifying applicable existing standards, this community might first move to determine small sets of high priority common data elements for CBK metadata. Then, using agile methods for rapid iteration, following standards development best practices, a CBK metadata standard supporting its “FAIR”ness may result.

REFERENCES

1. Friedman CP, Wong AK, Blumenthal D. Achieving a Nationwide Learning Health System. *Sci Transl Med*. 2010 Nov 10;2(57):57cm29-57cm29.
2. Hammond WE, Richesson RL. Standards Development and the Future of Research Data Sources, Interoperability, and Exchange. In: *Clinical Research Informatics* [Internet]. Springer; 2012. p. 335–365. Available from: http://link.springer.com/chapter/10.1007/978-1-84882-448-5_18
3. Hagstrom S. The FAIR Data Principles [Internet]. FORCE11. 2014 [cited 2017 Oct 4]. Available from: <https://www.force11.org/group/fairgroup/fairprinciples>
4. Hammond WE, Jaffe C, Cimino JJ, Huff SM. Standards in biomedical informatics. In: *Biomedical informatics* [Internet]. Springer; 2014. p. 211–253. Available from: http://link.springer.com/chapter/10.1007/978-1-4471-4474-8_7/fulltext.html
5. Richesson RL. Data standards in diabetes patient registries [Internet]. SAGE Publications; 2011. Available from: <http://journals.sagepub.com/doi/abs/10.1177/193229681100500302>
6. ANSI. ANSI Essential Requirements: Due process requirements for American National Standards. January; 2013.
7. DCMI: Using Dublin Core [Internet]. [cited 2017 Oct 5]. Available from: <http://www.dublincore.org/documents/usageguide/elements/>



Appendix Figure. *Draft Figure to generate Comments and Promote Discussion*

This draft figure shows potential Categories of *General* and *Instance-Specific* CBK Metadata supporting CBK "FAIR"ness. A variety of common data elements (CDEs) pertain. General metadata are thought to be needed for all instances of CBK. Some general metadata elements are shown in three broad categories of Type, Certification, and Provenance. A variety of other metadata elements, and their related metadata standards, may be specific to collections of CBK and not generally applied to all CBK. Two potential categories of instance-specific metadata may describe *Services* enabled by CBK and the *Domain(s)* to which the CBK pertains.