The Computable Biomedical Knowledge Metadata Model (CBK-MM) Project
Project Description, FINAL August 31, 2022

This project is a foundational and cross-community project for Mobilizing Computable Biomedical Knowledge (MCBK) that originated in the MCBK Technical Standards and Infrastructure (TSI) Work Group and will benefit greatly from the input and engagement of experts and viewpoints from the MCBK Sustainability & Inclusion workgroup, as well as the governance and policy expertise from members of the MCBK Trust & Policy workgroup. *The purpose of this document is to provide a high-level overview and non-technical overview of the CBK-MM project that can be used to facilitate an understanding of the project and the participation of all MCBK members. As the vision and strategy for CBK-MM is still evolving, this document will continue to be updated.*

_____

The vision of the CBK-MM project is to enable trustworthy creation and use of computable knowledge metadata for the web by developing a vocabulary and structure to represent knowledge objects, artifacts, or resources[1] so that they can be mobilized for action in biomedicine. The CBK-MM is to be a *semantic metadata model* that can support the sharing of computable knowledge resources across the internet. The CBK-MM model can be used by: 1.) **Knowledge creators** to describe their own CBK to make it understandable, 2) **Knowledge publishers** to index CBK to make it more findable, 3) **Knowledge managers** (stewards and librarians) to assure that CBK can be found, accessed, and used by others, and 4) **Knowledge users** who need to find, access, implement and re-use existing knowledge sources. Enabled with the CBK-MM shared open vocabulary, individuals or organizations (both public and commercial) can make their own "knowledge catalogs" by collecting knowledge resources based on what is published, pooling information across catalogs, and integrating knowledge based on that published metadata. Additionally, a CBK-MM could enable **knowledge curators / auditors** to assemble, assess, audit, and curate collections of CBK that are current, relevant, and trusted for specific user communities and purposes. Further, the use of an open vocabulary like CBK-MM will allow sustainable publication of agreed upon computable knowledge metadata for the internet.

Guiding Principles for CBK-MM Design and Development:

CBK-MM will serve as an interchange language for computable knowledge resource metadata on the web using Linked Data principles (optimized for JSON-LD like Schema.org).

We will keep in mind the following guiding principles when developing CBK-MM:…

- Focus on enabling CBK FAIR+T (findable, accessible, interoperable, re-usable, and trustable) for the CBK metadata, and will prioritize "open" CBK
- Re-use existing standards (i.e., refrain/minimize the development of new standards)
- Employ a minimal number of elements (i.e., simple)
- Apply to ALL CBK types so to promote generalizability
- Make extensible to allow detailed metadata or ontologies to represent CBK-specific features for different types of CBK

_____

[1] We consider knowledge object, knowledge artifact, and knowledge resource to be somewhat interchangeable as terms, and use them interchangeably in this document.

- Employ a flexible governance structure that promotes sustainability and inclusivity

Strategy and Starting Points for CBK-MM:

The strategy for CBK-MM is to identify *minimum* metadata that can describe knowledge objects. The vision and goal of CBK-MM are to enable CBK and knowledge objects to be FAIR + T (findable, accessible, interoperable, re-usable, and trustable).

CBK-MM will be developed to support the following broad use cases for the browsing, search, and retrieval of interoperable CBK / knowledge objects. Consumers of CBK-MM metadata will be able to:

- Index and search human-readable descriptions and labels of CBK resources.
- Filter CBK resources by representation, access method, terms of use, and topic to find resources suitable for their application.
- Use the metadata to determine access methods to automatically integrate knowledge when possible.

The W3C's Dataset Catalog (DCat) is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. As a starting point, the CBK-MM group will look at the DCat standard and development strategy to see if we can develop / extend / augment it to represent knowledge (rather than data sets).

The group will also consider leveraging the Dublin Core Metadata Terms (DC Terms), Schema.org , and the W3C Provenance Ontology (PROV-O), which defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web. We are also aware of possibly relevant efforts from the Research Data Alliance (RDA) working groups around crosswalks between different schema for metadata around data sets and will reach out to these communities to better understand how they may be leveraged here.

Our group is aware of and hopes to include (not rebuild) the previous MCBK Standards work including various use cases for CBK and a typology of CBK, as well as the definition of broad categories of CBK found in Alper et al (2020): https://onlinelibrary.wiley.com/doi/full/10.1002/lrh2.10271 The group is also aware and supports the Common Metadata Framework project which is developing mappings across multiple existing metadata schema. Additionally, we want to incorporate the work of other MCBK workgroups, in particular the library perspective and bias work from the Sustainability & Inclusion group and the trust attributes of knowledge objects developed by the Trust & Policy working group.

Development, Collaboration, and Governance of CBK-MM:

CBK-MM is now a project of the MCBK TSI Work Group and includes liaisons from the Sustainability & Inclusion and Trust & Policy MCBK workgroups.

All MCBK members are welcome to join the meetings and contribute to the development of the vision and the standards.

Following W3C working group methods, we will plan for the following:

- Establish and plan consensus-based process to advance goals of the project
- Invite all stakeholders to join and contribute to consensus-based development
- Accept, discuss, and approve proposals for expanding the metadata model
- Vote on milestone goals for releases
- Write documentation and working examples

A plan for future governance and sustainability of CBK-MM is needed. We look to the MCBK Trust & Policy and Sustainability & Inclusion Working Groups for guidance on that.