Version of June 23, 2020

Developing Metadata Categories as a Strategy to Mobilize Computable Biomedical Knowledge

Brian S. Alper	*EBSCO Information Services and University of Missou*ri
Bruce E Bray	*University of Utah*
Marisa L. Conte	*University of Michigan*
Christina Eldredge	*University of South Florida*
Allen Flynn	*University of Michigan*
Sigfried Gold	*University of Maryland*
Robert A. Greenes	*Arizona State University and Mayo Clinic*
Peter Haug	*Intermountain Healthcare and University of Utah*
Gunes Koru	*University of Maryland*
James McClay	University of Nebraska
Marc L. Sainvil	*Mayo Clinic*
Davide Sottara	*Mayo Clinic*
Mark Tuttle	*Apelon*
Robin Ann Yurk	*MDyurk*

INTRODUCTION

Computable biomedical knowledge artifacts (CBKs) are digital objects or entities representing biomedical knowledge as machine-independent data structures that can be parsed and processed by different information systems. The breadth of content represented in CBKs spans all biomedical knowledge related to human health and so it includes knowledge about molecules, cells, organs, individual people, human populations, and the environment.

CBKs vary in their scope, purpose, and audience. Some CBKs support biomedical research. Other CBKs help improve health outcomes by enabling clinical decision support, health education, health promotion, and population health analytics. In some instances, CBKs have multiple uses that span research, education, clinical care, or population health. As the number of CBKs grows large, producers must describe them with structured, searchable metadata so that consumers can find, deploy, and use them properly. This report delineates categories of metadata for describing CBKs sufficiently to enable CBKs to be mobilized for various purposes.

Different types of CBKs exist. CBK types include value sets[1], terminologies and ontologies[2,3], computable phenotypes[4], computable recommendations from guidelines[5], computable evidence resources[6], predictive models[7], causal models[8], and business process models[9].

CBKs produced by data scientists and knowledge engineers are an increasingly common form of scholarly communication[10]. In some circumstances, people publish CBKs so they can be replicated, reproduced, and used by others[11]. Following the example set by journals in computer science, biomedical journals are beginning to support the publication of the computer-processable artifacts that are used for or produced by scientific studies[12].

We believe CBKs are essential for large-scale Precision Health[13] and critical to the success of rapid Learning Health Systems[14]. Precision Health and Learning Health System initiatives and computation-intensive. To be effective, Precision Health and Learning Health Systems rely on the systematic application of complex CBKs to bring about the widespread gains needed to meet the Quintuple Aim[15].

Moreover, students and clinical educators are important CBK stakeholders. As curricula in biomedicine evolve, we anticipate more students will develop and use CBKs during their training for careers in biomedical science, the health professions, or related disciplines[16].

To describe CBKs to more easily mobilize them for research, education, clinical care, and public health, we identified key categories of **CBK metadata**. We focused on CBK metadata that uphold trust and make CBKs findable, accessible, interoperable, and reusable (the FAIR principles)[17].

BACKGROUND AND SIGNIFICANCE

About Computable Biomedical Knowledge Artifacts (CBKs)

All CBKs are Digital Objects. Work on Digital Object metadata predates Kahn and Wilensky's 1995 Framework for Distributed Digital Object Services[18]. The three fundamental components of all Digital Objects are **content** (in the form of a bit sequence), a unique **identifier**, and describable **properties** (e.g., size in bits)[19]. This work explores which describable properties are useful to mobilize CBKs.

CBKs often get incorporated into larger software applications in ways that make them difficult to identify, isolate, extract, and share[20]. However, we assume that all CBKs can be isolated and shared as independent digital objects, depending on software design[21,22]. Therefore, we do **not** consider applications (apps) or software services (APIs) that incorporate CBKs to be CBKs. CBKs are digital objects that represent biomedical knowledge in machine-independent data structures. They may either stand alone or be embedded within apps or APIs.

We are **not** aware of any comprehensive CBK typology. Instead, in this work on CBK metadata, we draw on two different perspectives about CBK types. First, CBK types may reflect the structured computer-processable **formats or languages** used to represent their knowledge content (e.g., propositional logic or Python)[23]. Second, CBKs may be distinguished by their place in a hierarchy of increasing CBK **complexity**[24]. The hierarchy we have in mind builds on basic CBKs like **terms** and **relationships** to arrive at increasingly complex composite CBKs like **decision trees**, **workflows**, and **plans** [20]

Regardless of type, all CBKs carry knowledge content about biomedicine. Per Friedman, knowledge "is the result of an analytic or deliberative process that holds significance for an identified community"[25]. Per Newell and Simon, the significance of knowledge arises from the role it plays in assisting communities in achieving their goals[26]. We value CBKs because they can help people and communities achieve important scientific and health goals.

Summarizing, we view CBKs as Digital Objects that are concrete, distinct, shareable material *information content entities* [27,28]. CBKs represent and convey biomedical evidence or results when they hold significance for an identified community and are explicitly formatted to enable their immediate processing or execution by digital computers.

About Mobilizing CBKs

The Mobilizing Computable Biomedical Knowledge (MCBK) movement calls for the development of open, safe, effective, equitable, and inclusive CBKs that can be trusted[29] . Once suitable CBKs are developed, metadata are needed to communicate their many and various attributes to CBK users [30].

Public and private repositories of CBKs are under development. CDS Connect[31] and the Value Set Authority Center[32] are two public repositories. Other examples include the computable phenotype repository PheKB[4], the Kipoi repository of predictive models for genomics[7], and the DDMORE repository of computable models for pharmaceutics[33]. Some suggest that private software code repositories, such as GitHub, Souceforge, and Bitbucket, are suitable for hosting CBKs [34]. However, others point out that the policies governing these repositories may not fully support the CBK long-term sharing needs of biomedical scientists [35,36].

Many prior works on metadata guide this effort to develop a strategy to mobilize CBKs. The most notable prior works focus on making **data sets** FAIR[17]. Globally, there are many efforts ongoing to develop and promote metadata standards to annotate data sets. Organizations and projects such as the FORCE11[37], CEDAR[38,39], GO FAIR[40], DataCite[41], and the Research Data Alliance[42], are advancing support for metadata about data sets. We build on what is already being done for **data set** metadata to develop metadata categories for CBKs.

The barriers to creating sufficient metadata to mobilize CBKs by making them trustable and FAIR are high [38]. Producers of CBKs need to have substantial resources to author metadata manually or generate metadata automatically. Consequently, CBK producers need ways to minimize and recoup the costs of providing sufficient metadata.

Consumers also have a role to play in creating and providing metadata about CBKs. For example, consumers of CBKs can provide feedback about outcomes and issues that arise from using CBKs in practice. We believe metadata about the real-world deployment, integration, application, and use of CBKs is important to uphold trust[43], making these metadata critical for any large CBK ecosystem's success.

We anticipate that the production of CBKs will continue to increase as it has since the 1970s[44]. Mobilizing the growing number of CBKs to achieve optimal use requires them to be properly organized and managed. This work significantly advances a strategy to mobilize CBKs by specifying and using CBK metadata. Further development of metadata for CBKs can enable them to be widely shared and appropriately used for initiatives that focus on research, education, health promotion, health care, population health, and public health.

RESEARCH QUESTION

Which existing or new **categories of metadata elements** hold the potential to help make CBKs trustable, findable, accessible, interoperable, and reusable?

METHODS

As a strategy to mobilize CBKs, researchers, data scientists, and knowledge engineers who are members of the MCBK movement's Standards Workgroup (SWG) came together to develop a profile of CBK metadata categories. We performed this research in 2020 during weekly videoconferences and other small group meetings. Three phases of development led to the development of our initial metadata profile for CBKs: 1) performing an environmental scan, 2) surfacing candidate metadata categories, 3) deciding upon an initial CBK metadata profile.

Phase 1 – Environmental Scan

In February of 2020, we conducted a rapid environmental scan of metadata specified by existing standards and used to describe CBKs held in existing repositories. We reviewed metadata and metadata categories from Health Level 7, Dublin Core, Schema.org, OMG.org, GitHub, FORCE11, and the Library of Congress. We compiled information about metadata elements from these sources into a spreadsheet for analysis.

Phase 2 – Surfacing Candidate Metadata Categories

We iteratively analyzed potential metadata categories by applying an evolving list of categories to several real-world examples of CBKs with metadata available in existing repositories. After several cycles of applying, discussing, and refining our categories list, we realized 15 candidate metadata categories for an initial draft of our CBK metadata profile.

For each candidate category that we surfaced this way, we listed specific elements included in the category and then tried to identify prior works that supported including each candidate category in our final metadata profile. During this phase, we also attempted to identify how metadata elements in each candidate category related to making CBKs trustable and FAIR. A supplement at the end of this document provides examples showing actual metadata for several different types of CBKs.

Phase 3 – Deciding Upon an Initial Metadata Profile

When deciding on which metadata categories to keep and which to combine or set aside, we gave preference to previously defined metadata categories over new categories. As part of our decision-making process, we clarified the scope of the metadata categories in our initial metadata profile by drafting and revising a paragraph outlining each category's scope. Finally, once our group had decided upon an initial metadata profile, we examined the profile to create a research agenda focused on requirements for metadata that make CBKs trustable and FAIR.

RESULTS

Profile of Metadata Categories

The main result of this work is a profile of metadata categories for describing CBKs. The profile specifies eleven categories of metadata elements (Table 1). A description of each metadata category in the profile comes next.

1. Biomedical Domain Metadata – [Example: *has domain of human genetics*.] Biomedical Domain Metadata serve as tags to help ensure that CBKs can be found via search with high recall and precision. Support for generating these metadata comes from many sources, such as the Unified Medical Language System (UMLS). By processing information about CBKs, tools like MetaMap[45] have the potential to generate Biomedical Domain Metadata automatically. Like keywords for journal articles, elements describe relevant biomedical domains for CBKs with varying degrees of specificity (e.g., cardiology vs. glycogen synthase kinase). Biomedical Domain Metadata support the internationalization of CBK search via translation of tags.

2. Coverage Metadata – [Examples: 1. *from a study of adult women between 70 and 80 years of age* 2. *from a study of immortalized human hepatic cells*.] Coverage Metadata have several sub-dimensions [46]. Elements circumscribe relevant CBK biological samples or human populations, situations, temporalities, and limitations. Coverage Metadata convey aspects of generalizability, including potential and actual biases. These metadata support search and findability, enable appropriate reusability, and support trust.

3. Integrity Metadata – [Example: *has SHA1 hash of D7D095AD616FF051AD41F3935242537D7DDAF0C9*] CBKs can be distributed over computer networks. In network environments, integrity metadata are used by senders and receivers to verify CBK authenticity and completeness. Cryptographic hash functions provide a mechanism that allows fetched CBKs to be checked for tampering in transit. For the most part, integrity metadata are processed by machines and not by people. An existing specification for integrity metadata is available from the W3C. Integrity Metadata elements prevent unwarranted manipulation of CBKs, and thus they directly support trust in CBKs.

4. Performance Metadata – [Example: *has an average rating of 4 out of 5 stars from 100 verified users*]. Performance Metadata elements describe records of CBK use and impacts from the CBK consumer's point of view. Elements include information about validation studies and other test results, user assessments or ratings, user-assigned metrics, measures, or grades, and any specific outcomes from CBK deployment and use that are of interest to CBK consumers. These metadata directly support reusability and trust. They also support search and findability by user or by metric.

METADATA CATEGORY	METADATA ELEMENT COVERAGE BY CATEGORY	ELEMENT FUNCTIONS AND USES	BASED ON
1. Biomedical Domain	Elements describing **biomedical domains** to which CBKs are related	• Convey relevance • Enable cross-terminology mappings • Support internationalization • Support search and findability	1, 2
2. Coverage	Elements circumscribing CBK population, spatial, or temporal applicability including descriptions of **cohorts** and **audiences**, **inclusion-exclusion criteria**, **situations** and **contexts** for use, **periods of use**, **biases** or other **limits on generalizability**	• Enable reusability • Support search and findability • Support trust	3, 4
3. Integrity	Elements conveying **outputs from cryptographic functions** that allow CBK users to confirm that a CBK has been delivered without tampering	• Prevent unwarranted manipulation • Support trust	5
4. Performance	Elements that describe CBK use from a consumer point of view in terms of **performance metrics**, **test results, validation studies,** or **outcomes from use**	• Support reusability • Support search and findability • Support trust	6
5. Preservation	Elements needed specifically to **archive** CBKs for long periods of time without any degradation	• Enable future research • Support accessibility • Support digital forensics • Support root cause analyses • Support trust	7, 8
6. Provenance	Elements that describe **changes in ownership**, **custody**, **composition, or version** that take place throughout CBK lifecycles	• Convey lifecycle events with dates • Convey ownership • Enable publishing • Enable versioning • Support search and findability • Support trust	9
7. Purpose	Elements indicating **intended uses** for CBKs and describing the **motivations of CBK producers** and the **goals producers have** for CBKs	• Convey relevance • Support search and findability	
8. Rights Management	Elements to describe sources of any **legal rights** reserved or foregone by owners of CBKs	• Support reusability	9
9. Source	Elements indicating **sources** from which CBKs spring, **knowledge resources** to which CBKs relate**, repositories** where CBKs are stored, or any elements used to **identify** and **enable access** to CBKs	• Convey attribution • Distinguish among CBKs • Support search and findability • Support accessibility • Support trust	4, 11, 12
10. Technical	Elements to describe a wide array of **technical features** of CBKs that are specifically needed to **deploy, integrate, and operate** them	• Enable deployment and execution • Support interoperability • Support search and findability	3, 9
11. Type	Elements used to classify CBKs **by format**, **structure**, **methods of creation**, etc.	• Support accessibility • Support interoperability • Support search and findability	

1. Chong Q, Marwadi A, Supekar K, Lee Y. Ontology-based metadata management in medical domains. Journal of Research and Practice in Information Technology.2003 May;35(2):139.
2. Buendía F, Gayoso-Cabada J, Juanes-Méndez JA, Sierra JL. Transforming Unstructured Clinical Free-Text Corpora into Reconfigurable Medical Digital Collections. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) 2019 Jun 5 (pp. 519-522). IEEE.
3. Dublin Core Metadata Initiative. Dublin core metadata element set, version 1.1.
4. Lehmann HP, Downs SM. Desiderata for sharable computable biomedical knowledge for learning health systems. Learning health systems. 2018 Oct;2(4):e10065.
5. W3C. Integrity Metadata. At https://www.w3.org/TR/SRI/#integrity-metadata accessed May 3, 2020.
6. Wroe C, Goble C, Greenwood M, Lord P, Miles S, Papay J, Payne T, Moreau L. Automating experiments using semantic data in a bioinformatics grid. IEEE Intelligent Systems. 2004 Jan;19(1):48-55.
7. Caplan P. Understanding PREMIS. Washington DC, USA: Library of Congress.
8. Miksa T, Rauber A, Mina E. Identifying impact of software dependencies on replicability of biomedical workflows. Journal of biomedical informatics. 2016 Dec 1;64:232-54.
9. Lebo T, Sahoo S, McGuinness D, Belhajjame K, Cheney J, Corsar D, Garijo D, Soiland-Reyes S, Zednik S, Zhao J. Prov-o: The prov ontology. W3C recommendation. 2013 Apr 30;30.
10. Beyene WM, Godwin T. Accessible search and the role of metadata. Library Hi Tech. 2018 Mar 19.
11. Kunze J, Baker T. The Dublin core metadata element set. RFC 5013, August; 2007 Aug.
12. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3.

Table 1. Profile of metadata to make CBKs trustable and FAIR.

5. Preservation Metadata – [Example: *has preservation level of high, meaning a minimum of five independent copies are kept on a variety of storage media*] Preservation Metadata represent the information needed for the conservation of CBKs over long durations measured in decades. Elements support long-term archiving by indicating things like the planned duration of archiving and by specifying various methods of digital preservation. These metadata play a special role by supporting root cause analyses of incidents involving CBKs after CBKs have been taken out of use. These metadata also support the safekeeping of CBKs for future research. In these ways, Preservation Metadata facilitate accessibility and uphold trust through verification.

6. Provenance Metadata – [Example: *created on June 16, 2016 by ACME corporation*] Provenance Metadata record key events in CBK lifecycles, including changes in ownership, custody, or composition. These metadata directly support a variety of CBK versioning methods for CBKs. Elements may be fine- or coarse-grained depending on the level of detail needed about the lifecycles of CBKs. These metadata support searching for CBKs by lifecycle event, lifecycle event dates, versions or lifecycle stages. Provenance Metadata uphold trust by providing a mechanism to track and trace CBKs from their origin through to their use in practice and ultimate withdrawal and replacement.

7. Purpose Metadata – [Example: *is designed for cohort identification*] Purpose Metadata describe reasons why a CBK was designed and how its producers intend for it to be used. Elements describe **motivations** and **goals** for CBKs from the producer's perspective. To convey purpose, these metadata utilize predicates such as '*is_designed_to'*, '*is_intended_to'*, '*is_not_designed_to'*, or '*is_not_intended_to'*. These metadata convey relevance, support search and findability, and guide consumers to use CBKs as intended.

8. Rights Management Metadata – [Example: *has license Creative Commons ND 2.0*] Rights Management Metadata primarily stipulate legal information governing the use and reuse of CBKs. Elements may specify licenses, copyrights, disclaimers, warranties, or legal limits on use. These metadata support reusability. They also enable searching by license or other legal policies.

9. Source Metadata – [Example: *has original source identified by PMID:19762550*.] Source Metadata are a diverse category of metadata supporting CBK access and management. Elements include organizational sources of CBKs, resources upon which CBKs are founded, attributions for CBKs, unique CBK identifiers, repositories that store CBKs, and CBK locations. Source Metadata uniquely identify CBKs for users. These metadata primarily support search, findability, accessibility, and trust.

10. Technical Metadata – [Example: *has file format of .bpm*] Technical Metadata describe a wide array of CBK technical features supporting consumers who wish to deploy, integrate, and operate CBKs effectively. These metadata directly support interoperability and enable search and findability according to salient technical details.

11. Type Metadata – [Example: *has CBK artifact type of value set*] Type Metadata are used to group and classify CBKs by certain features of interest. Elements may convey types in a general way (e.g., predictive model, causal model, order set), or they may convey specific CBK subtypes (e.g., Disease Computable Phenotype vs. Drug Response Computable Phenotype). The metadata functions support accessibility, interoperability, search, and findability.

Research Agenda for CBK Metadata

Our work to create the CBK metadata profile above revealed a number of CBK metadata requirements requiring further analysis or broader examination. Seven of these requirements comprise the research agenda appearing in Table 2 below.

Overall, we recognize that a lot of additional work is needed to define and organize the metadata elements in each category of the CBK metadata profile. This is especially true for the purpose and type categories, which both have elements that are not yet sufficiently defined.

RESEARCH AGENDA ITEM	BRIEF DESCRIPTION OF RESEARCH AGENDA ITEM	RELATED METADATA CATEGORY
Requirement 1 - **Standardize Descriptions of Relevant Samples and Populations**	A single CBK may be related to a variety of biological samples or human populations. Examples include proteins and other molecules, cell lines, or cohorts of people studied to generate CBKs. Standards for metadata describing relevant samples and populations need to be developed or identified and then adopted.	Coverage
Requirement 2 - **Standardize Descriptions of CBK Biases**	Many biases may pertain to CBKs. Some biases may be quantifiable. Cases of CBK biases continue to raise concerns. Standards for metadata describing certain biases already exist. More work is needed to identify and adopt standards for metadata that describe CBK biases.	Coverage
Requirement 3 - **Define CBK Use Outcomes of Interest**	It is not clear which outcomes from using CBKs are of most interest to consumers. Studies of CBK user needs for Performance Metadata by CBK type can lead to a better understanding of the outcomes of greatest interest to CBK consumers.	Performance, Type
Requirement 4 - **Define and Describe CBK Lifecycles**	The lifecycles of CBKs need to be better understood. Since CBK lifecycles may vary by CBK type, interactions between Provenance Metadata and Type Metadata needs to be explored.	Provenance, Type
Requirement 5 - **Develop Schema for Purpose Metadata**	Purpose Metadata to convey producer goals, motivation, and intent are evidently not well established. For this reason, work is needed to develop and trial new schemas for CBK Purpose Metadata.	Purpose
Requirement 6 - **Standardize Descriptions of Technical Metadata**	A single CBK may have a lot of technical characteristics. Many technical characteristics already appear in other metadata schemas. Work is needed to identify and adopt standards for metadata that describe the technical details of CBKs.	Technical, Type
Requirement 7 - **Develop Scrupulous CBK Typologies**	A variety of different approaches have been taken to define the types and subtypes of CBKs. Work is needed to synthesize these efforts into coherent typologies before standards for metadata about CBK types can be developed.	Type

Table 2. Research Agenda for CBK Metadata Analysis and Study

DISCUSSION

We envision a future state when CBKs will be widely shared to support biomedical research, education, and health improvement. Members of the MCBK community are working to develop standards and technical platforms for managing CBKs on a large scale. So that CBKs can be easily found and appropriately used, the MCBK community has prioritized work on CBK metadata standards. This effort to develop a CBK metadata profile is part of that work.

In our explorations, we found that many different types of CBKs already exist. By inspecting current metadata describing actual CBKs, we realized that one set of specific metadata elements will not suit to describe all types of CBKs. For this reason, we chose to specify **metadata categories** instead of individual metadata elements.

We present a metadata profile for CBKs composed of eleven new and existing categories of metadata elements. Our metadata profile specifies sufficient metadata to make CBKs trustable, findable, accessible, interoperable, and reusable. Most of the categories in the profile apply to data sets and other Digital Objects. However, some categories are unique to CBKs, e.g., Purpose Metadata and Type Metadata.

The scope of our metadata profile is large. It focuses on the metadata needs and contributions of two stakeholder groups, CBK producers and CBK consumers (or users).

We imagine that CBK producers will need to provide an as yet undefined minimum set of metadata to support CBK consumers. Besides providing fundamental metadata elements, such as persistent unique identifiers and locations where CBKs can be accessed, further work is needed to define the rest of the elements in the minimum necessary set of metadata users need from CBK producers.

CBK consumers will also need to provide metadata about CBKs for other users. For example, Performance Metadata describing the results of field testing and real-world use of CBK must come from those who use it after it has been produced.

Another important consideration is whether most CBK repositories will eventually span many CBK types or will tend to specialize by organizing one CBK type. If the evolution of **data set** repositories sets the pattern, then we can expect slow consolidation of tens of specialized CBK repositories into fewer, larger and more general CBK repositories.

Only two categories of metadata in the profile are seemingly new. The other categories in our profile have been previously described. The two new categories are CBK Purpose Metadata and CBK Type Metadata. We believe both of these categories need to be further analyzed and studied to arrive at workable frameworks, schema, and typologies for describing why CBKs exist and for adequately classifying CBKs.

Findability is supported by essentially all of the metadata categories in the profile. Using the specified categories of CBK metadata, researchers, clinicians, patients, and consumers can search by biomedical domain, source, and purpose to find CBKs. Software developers and

integrators can search by technical or type metadata to find CBKs that will operate in their information technology environments.

All categorization schemes are imperfect and incomplete[47]. We encountered several challenges while trying to develop a sufficient and useful metadata profile for CBKs. One challenge is to minimize overlap of metadata categories. As much as possible, we tried to identify distinct categories. Another challenge is to remain realistic about the number of metadata elements needed to make CBKs trustable and FAIR. We are concerned that the costs of generating and managing sufficient CBK metadata will be high, thus limiting widespread CBK sharing and use. For this reason, the value of every metadata element ultimately needs to be determined.

Standardization efforts are never complete. We look forward to learning more about CBK metadata from the real-world experiences of researchers, educators, clinicians and other consumers who use CBKs in their work. Despite the challenges, the metadata categories emerging from this study provide a profile that illuminates some CBK metadata needs. As a strategy to mobilize CBK, we look forward to further developing and refining our CBK metadata profile. An iterative approach can realize the goal of making CBKs trustable and FAIR.

CONCLUSION

CBKs vary widely in complexity, goals, and anticipated audience. Each CBK offers knowledge of potential value for clinical care, public health, education, or advancing biomedical science. Sharing is key to gain this value. To mobilize CBKs effectively, the value from sharing has to be greater than the costs of capturing and representing knowledge in computer-processable formats. For the producers of CBKs, dissemination to those able to benefit is of prime importance. For the consumers of CBKs, the ability to readily discover, deploy, and use CBKs to meet their clinical, educational, or scientific needs is most important.

Our metadata profile addresses the needs of CBK producers and consumers. It also enables CBK indexing and thereby promotes value, efficiency, dissemination, and discovery. We foresee a system that effectively indexes a very large number of CBKs. In this system, producers of CBKs publish salient CBK characteristics as metadata. Also, consumers can not only identify existing CBKs that match their criteria but also can determine the applicability of each CBK to the particular environment they are seeking to enrich.

Here, we have published and described our initial profile for a set of CBK metadata categories to serve the complementary needs of producers and consumers of computable biomedical knowledge. In addition, we have articulated a research agenda to guide further work towards having CBK metadata that is sufficient to uphold trust and to make CBKs FAIR.

ACKNOWLEDGMENTS

We thank Helen Pan for organizing and supporting our meetings. We thank Dr. Melissa Clarkson, Dr. Charles P. Friedman, and Dr. Rachel Richesson for their comments on early drafts of this paper. We are grateful for comments about CBKs received from Dr. Vojtech Huser.

References

1. Gold S, Batch A, McClure R, et al. Clinical Concept Value Sets and Interoperability in Health Data Analytics. In: *AMIA Annual Symposium Proceedings*. Vol 2018. American Medical Informatics Association; 2018:480.
2. Huff SM, Rocha RA, McDonald CJ, et al. Development of the logical observation identifier names and codes (LOINC) vocabulary. *J Am Med Inform Assoc*. 1998;5(3):276–292.
3. 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):1251–1255.
4. Kirby JC, Speltz P, Rasmussen LV, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *J Am Med Inform Assoc*. 2016;23(6):1046–1052.
5. Peleg M, Boxwala AA, Tu S, et al. The InterMed approach to sharable computer-interpretable guidelines: a review. *J Am Med Inform Assoc*. 2004;11(1):1–10.
6. Alper B, Mayer M, Shahin K, et al. Achieving evidence interoperability in the computer age: setting evidence on FHIR. *BMJ Evid-Based Med*. 2019;24(Suppl 1):A15.
7. Avsec Ž, Kreuzhuber R, Israeli J, et al. The Kipoi repository accelerates community exchange and reuse of predictive models for genomics. *Nat Biotechnol*. 2019;37(6):592–600.
8. Cooper G. Causal Network Discovery from Biomedical and Clinical Data. Published online 2018. Accessed May 13, 2020. http://hdl.handle.net/1853/59643
9. Müller R, Rogge-Solti A. BPMN for healthcare processes. In: *Proceedings of the 3rd Central-European Workshop on Services and Their Composition (ZEUS 2011), Karlsruhe, Germany*. Vol 1. ; 2011.
10. Hey T, Tansley S, Tolle K, others. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Vol 1. Microsoft research Redmond, WA; 2009.
11. Belhajjame K, Corcho O, Garijo D, et al. Workflow-Centric Research Objects: A First Class Citizen in the Scholarly Discourse. In: *SePublica@ ESWC*. ; 2012:1–12.
12. Wang W, Bleakley B, Ju C, et al. Aztec: A Platform to Render Biomedical Software Findable, Accessible, Interoperable, and Reusable. *ArXiv Prepr ArXiv170606087*. Published online 2017.
13. Williams MS, Buchanan AH, Davis FD, et al. Patient-centered precision health in a learning health care system: Geisinger's genomic medicine experience. *Health Aff (Millwood)*. 2018;37(5):757–764.
14. Etheredge LM. A rapid-learning health system: what would a rapid-learning health system look like, and how might we get there? *Health Aff (Millwood)*. 2007;26(Suppl1):w107–w118.
15. Matheny M, Israni ST, Ahmed M, Whicher D. Artificial intelligence in health care: The hope, the hype, the promise, the peril. *Natl Acad Med Prepub*. Published online 2020:94–97.
16. Stead WW, Searle JR, Fessler HE, Smith JW, Shortliffe EH. Biomedical informatics: changing what physicians need to know and how they learn. *Acad Med*. 2011;86(4):429–434.
17. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3(1). doi:10.1038/sdata.2016.18
18. Kahn R, Wilensky R. A framework for distributed digital object services. *Coporation Natl Res Initiat Rest*. Published online 1995.
19. Wittenburg P, Strawn G, Mons B, Boninho L, Schultes E. Digital objects as drivers towards convergence in data infrastructures (2019). *DOIURL Httpsdoi Org1023728b2share B605d85809ca45679b110719b6c6cb11*.
20. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a

systematic review. *Ann Intern Med*. 2012;157(1):29–43.
21. Flynn AJ, Shi W, Fischer R, Friedman CP. Digital knowledge objects and digital knowledge object clusters: unit holdings in a learning health system knowledge repository. In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE; 2016:3308–3317.
22. Mandl KD, Kohane IS, McFadden D, et al. Scalable collaborative infrastructure for a learning healthcare system (SCILHS): architecture. *J Am Med Inform Assoc*. 2014;21(4):615–620.
23. Boxwala AA, Rocha BH, Maviglia S, et al. A multi-layered framework for disseminating knowledge for computer-based decision support. *J Am Med Inform Assoc*. 2011;18(Supplement_1):i132–i139.
24. Fox J, Johns N, Rahmanzadeh A. Disseminating medical knowledge: the PROforma approach. *Artif Intell Med*. 1998;14(1-2):157–182.
25. Friedman CP, Flynn AJ. Computable knowledge: An imperative for Learning Health Systems. *Learn Health Syst*. 2019;3(4).
26. Simon HA, Newell A. Human problem solving: The state of the theory in 1970. *Am Psychol*. 1971;26(2):145.
27. Ceusters W, Smith B. Aboutness: Towards foundations for the information artifact ontology. Published online 2015.
28. Flynn AJ, Friedman CP, Boisvert P, Landis-Lewis Z, Lagoze C. The Knowledge Object Reference Ontology (KORO): a formalism to support management and sharing of computable biomedical knowledge for learning health systems. *Learn Health Syst*. 2018;2(2):e10054.
29. Manifesto of the Mobilizing Computable Biomedical Knowledge community movement. Accessed May 27, 2020. mobilizecbk.org
30. Lehmann HP, Downs SM. Desiderata for sharable computable biomedical knowledge for learning health systems. *Learn Health Syst*. 2018;2(4):e10065.
31. Lomotan EA, Meadows G, Michaels M, Michel JJ, Miller K. To Share is Human! Advancing Evidence into Practice through a National Repository of Interoperable Clinical Decision Support. *Appl Clin Inform*. 2020;11(01):112–121.
32. Bodenreider O, Nguyen D, Chiang P, et al. The NLM value set authority center. *Stud Health Technol Inform*. 2013;192:1224.
33. Harnisch L, Matthews I, Chard J, Karlsson M. Drug and disease model resources: a consortium to create standards and tools to enhance model-based drug development. *CPT Pharmacomet Syst Pharmacol*. 2013;2(3):1–3.
34. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. *PLoS Comput Biol*. 2013;9(10).
35. Shao H, Sun D, Wu J, et al. paper2repo: GitHub Repository Recommendation for Academic Papers. In: *Proceedings of The Web Conference 2020*. ; 2020:629–639.
36. Banks, Marcus. We Need a GitHub for Academic Research. *Slate*. Published online April 20, 2017. Accessed May 27, 2020. https://slate.com/technology/2017/04/we-need-a-github-for-academic-research.html
37. Rodriguez M. Research Communication Futures: A Perspective on the FORCE11 Scholarly Communication Institute. *Ser Rev*. 2018;44(4):307–312.
38. Musen MA, Bean CA, Cheung K-H, et al. The center for expanded data annotation and retrieval. *J Am Med Inform Assoc*. 2015;22(6):1148–1152.
39. Gonçalves RS, O'Connor MJ, Martínez-Romero M, et al. The CEDAR workbench: an ontology-assisted environment for authoring metadata that describe scientific experiments. In: *International Semantic Web Conference*. Springer; 2017:103–110.
40. Schultes E, Strawn G, Mons B. Ready, Set, GO FAIR: Accelerating Convergence to an

Internet of FAIR Data and Services. In: *DAMDID/RCDL*. ; 2018:19–23.
41. Starr J, Gastl A. isCitedBy: A metadata scheme for DataCite. Published online 2011.
42. Perez C. The RDA's Metadata Standards Directory: Information Gathering. Published online 2013.
43. Middleton B, Platt J, Richardson JE, Blumenfeld B. Recommendations for building and maintaining trust in clinical decision support knowledge artifacts. *Res Triangle Park NC Patient-Centered Clin Decis Support Learn Network2018*. Published online 2018.
44. Zhang L, Powell JJ, Baker DP. Exponential growth and the shifting global center of gravity of science production, 1900–2011. *Change Mag High Learn*. 2015;47(4):46–49.
45. Aronson AR. Metamap: Mapping text to the umls metathesaurus. *Bethesda MD NLM NIH DHHS*. 2006;1:26.
46. Kunze J, Baker T. *The Dublin Core Metadata Element Set*. RFC 5013, August; 2007.
47. Bowker GC, Star SL. *Sorting Things out: Classification and Its Consequences*. MIT press; 2000.

Supplement I – Examples of actual CBKs and their Metadata

The following supplement shows some of the work to find and evaluate existing metadata for actual CBKs. An **earlier version** of the metadata profile developed during this effort was used to develop these examples.

Calculator Artifact

https://www.mdcalc.com/cha2ds2-vasc-score-atrial-fibrillation-stroke-risk

Accessed: May 12, 2020

EXAMPLE METADATA for the CALCULATOR ARTIFACT AT THE LINK ABOVE:

METADATA CATEGORY	EXISTING METADATA
1. Access	URL: https://www.mdcalc.com/cha2ds2-vasc-score-atrial-fibrillation-stroke-risk
2. Applicability	Non-anticoagulated patient with non-valvular atrial fibrillation
3. Biomedical Domain	• The CHA2DS2-VASc score is one of several risk stratification schema that can help determine the 1 year risk of a thromboembolic • Stroke • Atrial Fibrillation • Risk Assessment
4. Integrity	Not present in the example
5. Performance	ORIGINAL STUDY: Validation study included 1,084 patients with non-valvular AF, not on anticoagulation, over age 18 with EKG or Holter diagnosed AF in the ambulatory and hospital settings from 182 hospitals in 35 countries from 2003 to 2004 and had known thromboembolic status at 1 year from the Euro Heart Survey database. VALIDATION STUDY: https://www.ncbi.nlm.nih.gov/pubmed/23408865 VALIDATION STUDY: https://www.ncbi.nlm.nih.gov/pubmed/22246443 VALIDATION STUDY: https://www.ncbi.nlm.nih.gov/pubmed/24759791 OUTCOMES STUDY: https://www.ncbi.nlm.nih.gov/pubmed/22922413
6. Preservation	Not present in the example
7. Provenance	CREATOR: Gregory Lip, MD OWNER: MDCALC.COM CONTRIBUTOR: Calvin Hwang, MD
8. Purpose	In 2012, the European Society of Cardiology (ESC) guidelines recommended a clinical practice shift, to initially focus on the identification of **'truly low risk'** patients who do not need any antithrombotic therapy. These low risk patients are those CHA2DS2-VASc score of 0 (male) or 1 (female). Subsequently, the next step is to offer effective stroke prevention (ie. Oral anticoagulation) to those with ≥1 additional stroke risk factors. Use the approach recommended in the 2012 ESC or NICE guidelines - first step, **identify LOW RISK patients**, i.e., CHA_2DS_2-VASc score of 0 (males) or 1 (females), who do not need any antithrombotic therapy, Next or subsequent step is to offer effective stroke prevention to all others with 1 or more additional stroke risk factors. As per the NICE guidelines, aspirin should not be used for stroke prevention in AF - it is minimally effective, not safe nor is it cost effective. CRTICAL ACTIONS: One recommendation suggests a 0 score is "low" risk and may not require anticoagulation; a 1 score is "low-moderate" risk and should consider antiplatelet or anticoagulation, and score 2 or greater is "moderate-high" risk and should otherwise be an anticoagulation candidate. • Consider not starting anticoagulation in patients with non-valvular AF and a CHA2DS2-VASc score of 0 as these patients had no TE events in the original study. • For those patients in whom anticoagulation is considered, risk bleeding scores such as ATRIA can be used to determine the risk for warfarin-associated hemorrhage. • Carefully consider all the risks and benefits prior to initiating anticoagulation in patients with non-valvular AF. • Some guidelines suggest that aspirin monotherapy is not supported by evidence.
9. Relation	ORIGINAL STUDY: CHA2DS2-VASc score (Birmingham 2009) was developed after identifying additional stroke risk factors in patients with atrial fibrillation.
10. Rights Management	COPYRIGHT: MDCalc All Rights Reserved TERMS OF USE: https://www.mdcalc.com/terms PRIVACY POLICY: https://www.mdcalc.com/privacy-policy
11. Technical	INPUTS: Age in years (categorical), Sex (binary), Congestive Heart Failure History (binary), Hypertension history (binary), Stroke history (binary), Vascular disease history (binary), Diabetes history (binary) OUTPUTS: (Score, Risk of ischemic stroke, Risk of stroke/tia/embolism), 0, 0.2%, 0.3%; 1, 0.6%, 0.9%; 2, 2.2%, 2.9%; 3, 3.2%, 4.6%; 4, 4.8%, 6.7%; 5, 7.2%, 10%; 6, 9.7%, 13.6%; 7, 11.2%, 15.7%; 8, 10.8%, 15.2%; 9, 12.2%, 17.4%
12. Type	ONLINE CALCULATOR – CALCULATOR

COMMENTS:

Citation Resource Artifact

http://gps.health/coka/resources/Citation/1

Accessed: May 12, 2020

EXAMPLE METADATA for the CITATION RESOURCE ARTIFACT AT THE LINK ABOVE:

METADATA CATEGORY	EXISTING METADATA
Access	Identifiers: COKA 1 (COVID-19 Knowledge Accelerator Citation #1), DOI 10.1101/2020.03.22.20040758, CORD UID q8l3ra55 url: https://gps.health/coka/resources/Citation/1
Biomedical Domain	Not present in the example, but classifier element can include MeSH terms or other biomedical domain classifiers
Applicability	Not present in the example, may not be relevant but classifier element can be used if needed
Integrity	Not present in the example, may not be relevant but classifier element can be used if needed
Performance	Not present in the example, may not be relevant but classifier element can be used if needed
Preservation	Version: 4 – this resource has been adjusted 3 times
Provenance	**Not present in the example, not built into the resource structure (but may be inherent in resource metadata in the FHIR resource structure)** Would "Source" fit here? authorString.source allows expression of the "source" for presentation of an author string articleUrl: https://www.medrxiv.org/content/10.1101/2020.03.22.20040758v2 authorString: Zhaowei Chen; Jijia Hu; Zongwei Zhang; Shan Jiang; Shoumeng Han; Dandan Yan; Ruhong Zhuang; Ben Hu; Zhan Zhang
Purpose	Inherent in the specification of ResourceType=Citation
Relation	**variantCitation: allows expression of relation to other Citation Resources (such as versions or parts of being cited)** **relatedIdentifier: allows expression of relation to other things without expression of type of relationship** **relatedArtifact: allows expression of relation to other things without expression of type of relationship** **medlinePubmed.relatedArticle: allows expression of relations to other articles with expression that relationship is determined by MEDLINE/PubMed**
Rights Management	**Copyright: used to express copyright for the abstract (the abstract may be the only copyrightable content in the Citation Reource)**
Technical	**publishingModel and journalIssue.citedMedium could code technical concepts in a high-level way**
Type	HL7 resourceType: Citation

COMMENTS: *Any Citation Resource artifact, like the artifact here, is intended to **describe** another knowledge artifact, the cited article. Thus, Citation Resource artifacts are essentially meta-knowledge artifacts that represent, carry, and convey computable biomedical knowledge **about** human readable knowledge (eg, knowledge about journal articles and books). As meta-knowledge artifacts, it is critical to view the Citation Resource artifact in its own right and not to confuse the Citation Resource artifact with the human-readable knowledge artifact it describes.*

Clinical Decision Support Artifact

https://cds.ahrq.gov/cdsconnect/artifact/screening-and-interventions-unhealthy-alcohol-use-logic-best-practice-alerts

Accessed: May 12, 2020

EXAMPLE METADATA for the CLINICAL DECISION SUPPORT ARTIFACT AT THE LINK ABOVE:

METADATA CATEGORY	EXISTING METADATA
1. Access	URL: https://cds.ahrq.gov/cdsconnect/artifact/screening-and-interventions-unhealthy-alcohol-use-logic-best-practice-alerts TITLE: Screening and Interventions for Unhealthy Alcohol Use: Logic for Best Practice Alerts
2. Applicability	This artifact is intended for use in an adult population aged 18 and older without a history of AUD INCLUSION: Populations for which alcohol use screening is desirable EXCLUSIONS: people who are terminally ill, people with known alcohol use disorder
3. Biomedical Domain	KEYWORDS: Unhealthy Alcohol Use, Screening, Brief Counseling, Intervention, Referral, MeSH TOPICS: Alcohol Drinking, Primary Health Care, Internal Medicine
4. Integrity	Not present in the example
5. Performance	IMPLEMENTATION REPORT: https://doi.org/10.23970/AHRQEPCMETHENGAGEALCOHOLGUIDE METHODS REPORT: https://doi.org/10.23970/AHRQEPCMETHENGAGEALCOHOL
6. Preservation	Not present in the example
7. Provenance	CREATE DATE: Oct 31, 2018 VERSION: 0.1 CREATOR: OWNER: **** PUBLISHER: RTI-UNC Evidence-based Practice Center STEWARD: Agency for Healthcare Research and Quality CONTRIBUTORS: Dan Jonas, Colleen Barclay, Shana Ratner, Julia Tompkins
8. Purpose	Intended for use by primary care clinical/medical directors in position to implement screening for unhealthy alcohol use The logic diagrams are intended to guide practices and health systems in creating electronic health record tools to facilitate screening and interventions for unhealthy alcohol use in primary care practices
9. Relation	REPOSITORY: CDS Connect REPOSITORY APPROVAL DATE: Dec 27 2018 RELATED CDS ARTIFACT: https://cds.ahrq.gov/cdsconnect/artifact/interventions-unhealthy-alcohol-use-smartset-note-documentation-referrals-and SUPPORTING EVIDENCE: The 2013 evidenced-based source that this artifact was derived from was updated by the USPSTF in 2018. The updated recommendation is available here: https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/unhealthy-alcohol-use-in-adolescents-and-adults-screening-and-behavioral-counseling-interventions
10. Rights Management	LICENSE: Federal Government Unlimited Rights License LICENSE URL: https://cds.ahrq.gov/cdsconnect/license/federal-government-unlimited-rights-license
11. Technical	FORMAT: SEMi-STRUCTURED TEXT FORMAT: PSEUDOCODE WORKFLOW: Interventions and Actions @ https://cds.ahrq.gov/cdsconnect/artifact/screening-and-interventions-unhealthy-alcohol-use-logic-best-practice-alerts
12. Type	CDS CONNECT TYPE: Alert

COMMENTS:

Computable Phenotype Artifact

https://phekb.org/sites/phenotype/files/FH_eAlgorithm_Pseudocode_FullText_2016_1_3.pdf
Accessed: May 12, 2020

EXAMPLE METADATA for the COMPUTABLE PHENOTYPE ARTIFACT AT THE LINK ABOVE:

METADATA CATEGORY	EXISTING METADATA
1. Access	URL: https://phekb.org/sites/phenotype/files/FH_eAlgorithm_Pseudocode_FullText_2016_1_3.pdf ABOUT PAGE: https://phekb.org/phenotype/602 TITLE: Electronic Health Record-based Phenotyping Algorithm for Familial Hypercholesterolemia
2. Applicability	Not present in the example
3. Biomedical Domain	• Familial hypercholesterolemia (FH) • Risk for premature atherosclerotic cardiovascular disease
4. Integrity	Not present in the example
5. Performance	VALIDATION STUDY: Safarova MS, Liu H, Kullo IJ. Rapid identification of familial hypercholesterolemia from electronic health records: The SEARCH study. J Clin Lipidol. 2016;10:1230-1239.
6. Preservation	Not present in the example
7. Provenance	OWNER: Mayo Clinic CREATOR: Iftikhar Kullo CONTRIBUTORS: Adelaide Arruda-Olson, MD, PhD , Carin Smith Smith., Hongfang Liu, PhD , Majid Rastegar , Maya Safarova, MD, PhD , Parvathi Balachandran, MBBS , Saeed Mehrabi, Sunghwan Sohn, PhD , Xiao Fan, PhD , Yijing Cheng VERSION: "SEARCH eAlgorithm for FH version 2.0 from June 2016" LAST UPDATED: June 2016 STATUS: Final
8. Purpose	We provide a pseudocode to identify cases and controls for primary hypercholesterolemia and Familial hypercholesterolemia (FH). This EHR-based algorithm is intended to optimize screening and identification of patients with FH among individuals with severe hypercholesterolemia and therefore increase awareness, detection and control of FH.
9. Relation	REPOSITORY NAME: PheKB REPOSIOTRY URL: https://phekb.org PheKB RECORD CREATOR: Safarova MS PheKB RECORD CONTRIBUTORS: Safarova MS, Liu H, Arruda-Olson A, Rastegar M, Smith C, Cheng Y, Fan X, Balachandran P, Sohn S, Kullo IJ
10. Rights Management	Not present in the example
11. Technical	FORMAT: PSEUDOCODE OUTPUT: Final output consists of (i) a case/control/unknown status for primary hypercholesterolemia, (ii) demographics of each individual (age at the time of qualifying LDL-C ascertainment, gender, race/ethnicity), (iii) lipid profile (total cholesterol, LDL-C, HDL-C, triglycerides), (iv) lipid-lowering treatment and difference in time between the index date and date of treatment ascertainment, (v) personal history of premature ASCVD and/or hypercholesterolemia, (vi) family history of premature ASCVD, (vii) xanthomas and/or early corneal arcus, (viii) Dutch Lipid Clinic Network score and case/control/unknown for FH status. CODING SYSTEMS USED: CPT Codes, ICD-9 Codes
12. Type	COMPUTABLE PHENOTYPE PheKB Type: Disease or Syndrome Computable Phenotype

COMMENTS:

Ontology Artifact

https://raw.githubusercontent.com/arpcard/aro/master/aro.owl
Accessed: May 12, 2020

EXAMPLE METADATA for the ONTOLOGY ARTIFACT AT THE LINK ABOVE:

METADATA CATEGORY	EXISTING METADATA
1. Access	URL: https://raw.githubusercontent.com/arpcard/aro/master/aro.owl NAME: Antibiotic Resistance Ontology ACRONYM: ARO ABOUT PAGE: https://github.com/arpcard/aro CONTACT: card@mcmaster.ca
2. Applicability	Not present in the example
3. Biomedical Domain	• Antibiotics • Drugs • Medications • Antibiotic Resistance
4. Integrity	Not present in the example
5. Performance	Not present in the example
6. Preservation	Not present in the example
7. Provenance	OWNER: McMaster University, Ontario, Canada CONTRIBUTOR: raphenya (GITHUB) LAST UPDATED: March 2020
8. Purpose	The Antibiotic Resistance Ontology describes antibiotic resistance genes and mutations, their products, mechanisms, and associated phenotypes, as well as antibiotics and their molecular targets. The Antibiotic Resistance Ontology (ARO) organizes the information describing the ability of a microorganism to withstand the effects of an antibiotic
9. Relation	INTEGRATED WITH: ARO is the Comprehensive Antibiotic Resistance Database (https://card.mcmaster.ca), a curated resource containing high quality reference data on the molecular basis of antimicrobial resistance. EDIT VERSION: src/ontology/aro-edit.owl
10. Rights Management	LICENSE: Creative Commons CC-BY license version 4.0 COPYRIGHT: McMaster University DISCLAIMER AT: https://card.mcmaster.ca/about LIMITATION OF LIABILITY AT: https://card.mcmaster.ca/about
11. Technical	FILE TYPE: .owl Number of concepts: 4498 Ontology Terms
12. Type	ONTOLOGY

COMMENTS:

Value Set Artifact

https://vsac.nlm.nih.gov/valueset/2.16.840.1.113883.3.464.1003.103.12.1001/expansion/Latest

Accessed: April 8, 2020

EXAMPLE METADATA for the VALUE SET ARTIFACT AT THE LINK ABOVE:

METADATA CATEGORY	EXISTING METADATA
Access	Title: ***Diabetes: Hemoglobin A1c (HbA1c) Poor Control (> 9%)*** Identifier:(2.16.840.1.113883.3.464.1003.103.12.1001 Owner (Steward): National Committee for Quality Assurance (NCQA) Contributors: Version (Expansion Version): eCQM Update 2019-05-10
Biomedical Domain	Diabetes | type 1 diabetes mellitus | type 2 diabetes mellitus | maternal diabetes mellitus
Generalizability/**context?**	One of 26 value sets for a clinical quality measure: **Diabetes: Hemoglobin A1c (HbA1c) Poor Control (> 9%),** https://ecqi.healthit.gov/sites/default/files/ecqm/measures/CMS122v8.html **Used by: eQualityMeasure**
Integrity	Not present in the example
Performance	Software tests included: Code quality analysis method included:
Preservation	Maintained by the NLM Value Set Authority Center and regularly updated to reflect newer vocabulary versions
Provenance	**Number of commits:** **Number of branches:** **Number of releases:** 12 **Latest commit:** 2019-05-10
Purpose	ICD10CM, ICD9CM, SNOMED CT codes indicative of a diagnosis of diabetes **Purpose:** • **Clinical Focus:** This value set contains concepts that identify patients who have a diagnosis of diabetes. • **Data Element Scope:** This value set may use the Quality Data Model (QDM) category related to Diagnosis. • **Inclusion Criteria:** Includes only relevant concepts associated with identifying patients who have type I or type II diabetes. • **Exclusion Criteria:** Excludes patients who have gestational diabetes or steroid-induced diabetes.
9. Relation	BACKGROUND: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371254/
Rights Management	**License:** uncertain whether still copyrighted with NCQA? **Disclaimer:** **Copyright:**
Technical	**Generated using: created by enumeration/created by hierarchy** **Operated using:**
Type	**Type statement:** Value set stored on VSAC

COMMENTS:

Order Set Artifact

https://cds.ahrq.gov/cdsconnect/artifact/endocrinology-hypoglycemia-order-set

EXAMPLE METADATA for the ORDER SET ARTIFACT AT THE LINK ABOVE:

METADATA CATEGORY	EXISTING METADATA
1. Access	URL: https://cds.ahrq.gov/cdsconnect/artifact/endocrinology-hypoglycemia-order-set TITLE: Endocrinology: Hypoglycemia Order Set
2. Applicability	This artifact is intended for use in an adult population with an active problem of either diabetes mellitus, type 1 or type Applicability Setting: Emergency Department (ED) | inpatient hospital | urgent care | ambulatory care clinic
3. Biomedical Domain	KEYWORDS: Endocrinology, Hypoglycemia, Diabetes Mellitus, Emergency Treatment
4. Integrity	Not present in the example
5. Performance	IMPLEMENTATION REPORT: https://cds.ahrq.gov/sites/default/files/cds/artifact/966/CCWP_B3B33B45B66Hypogly.pdf METHODS REPORT: https://cds.ahrq.gov/sites/default/files/cds/artifact/966/20180621REVISED-CDSK_KVRpt_OS_B33Hypogly_508.pdf
6. Preservation	Not present in the example
7. Provenance	CREATE DATE: April 20th, 2018 VERSION: 1.0 CREATOR: Veterans Health Administration OWNER: PUBLISHER: Veterans Health Administration STEWARD: Veterans Health Administration CONTRIBUTORS: Leonard Pogach MD, Paul Conlin MD
8. Purpose	Intended for use by clinical providers for Care and Management of Diabetic PatientS
9. Relation	REPOSITORY: CDS Connect REPOSITORY APPROVAL DATE: March 5th, 2019 RELATED CDS ARTIFACT: Hypoglycemia Rule https://cds.ahrq.gov/cdsconnect/artifact/endocrinology-hypoglycemia-rule SUPPORTING EVIDENCE: http://care.diabetesjournals.org/content/diacare/suppl/2016/12/15/40.Supplement_1.DC1/DC_40_S1_final.pdf
10. Rights Management	LICENSE: Apache COPYRIGHTS: 2018 Veterans Health Administration, Department of Veterans Affairs. All rights reserved. Contributions from external parties are property of respective copyright holders.
11. Technical	FORMAT: Structured code that is interpretable by a computer (includes data elements, value sets, logic)
12. Type	CDS CONNECT TYPE: Order Set

COMMENTS: