

Structures and Relations of Knowledge Nodes: Exploring a Knowledge Network of Disease from Precision Medicine Research Publications

Jian Qin, Ning Zou
Syracuse University

Abstract

The vast amount of DNA sequence and protein data are being explored and linked to diseases as causative factors to support clinical and healthcare decision making. These developments in data-intensive biological sciences and clinical practices raised new questions for knowledge organization systems (KOS), and taxonomies in particular. Sitting at the center of these questions is the lagging of KOS's capabilities in responding to the rapidly changing and emerging biomedical and disease terms due to the static, hierarchical structures and disconnection with new disease data in traditional KOSs. This paper reports a pilot study that is designed to uncover and identify the types of knowledge nodes and relationships that can help generalize a framework or model for building a Knowledge Network of Disease, or the New Taxonomy envisaged by the National Academy of Science. This pilot study examined a sample of biomedical publications and drew a knowledge map to lay out the main knowledge nodes and their relationships. A preliminary framework for constructing the Knowledge Network of Disease is discussed.

Keywords: knowledge nodes; Knowledge Network of Disease; content analysis; knowledge modeling

Citation: Editor will add citation

Copyright: Copyright is held by the authors.

Acknowledgements: We thank Marcia L. Zeng for her valuable comments during the paper writing and revision.

Contact: jqin@syr.edu

1 Introduction

Advances in DNA-sequencing technology have greatly impacted biomedical discoveries and turned biology into a data-intensive science. The vast amount of DNA sequence and protein data produced from biological research are being explored and linked to diseases as causative factors in support of clinical and healthcare decision making. These developments in data-intensive biological sciences and clinical practices raised new questions for knowledge organization systems (KOS), and taxonomies in particular. Sitting at the center of these questions is KOS's capabilities in responding to the rapidly changing and emerging biomedical and disease terms from data-intensive research. As the National Academy of Sciences (NAS) Committee on a Framework for Developing a New Taxonomy of Disease (the Committee thereafter) states it, "Because new information and concepts from biomedical research cannot be optimally incorporated into the disease taxonomy of today, opportunities to define diseases more precisely and to inform health-care decisions are being missed" (NAS, 2011, p. 3).

To address the problem of traditional taxonomy in the era of data-intensive biological sciences, the Committee envisions an "Information Commons" and a "Knowledge Network of Disease", in which biomedical data repositories as essential infrastructure will allow basic biological knowledge to be integrated with medical histories and health outcomes of individual patients for targeted, more precise medical decisions making. One of the recommendations from the Committee is to integrate biomedical research data into a disease knowledge network (NAS, 2011).

Traditionally, taxonomies such as the International Classification of Diseases (ICD) and the NCBI Taxonomy are structured hierarchically and parent-child relationships are dominant between classes in these taxonomies. Although various techniques, e.g., faceted classification and cross references, have been deployed to improve the relatedness between relevant classes, there are a number of obstacles that can prevent traditional taxonomies from integrating new knowledge generated from biomedical research data:

- The variations in coverage and scope of classifications created gaps currently existing between the taxonomies for biomedical research data and those for clinical practice. Clinical terms are in a very different domain from those used in biomedical science research, as such the consistency and effectiveness in communicating about diseases is affected between biomedical researchers and healthcare providers and administrators. This is reflected in the nature of current taxonomies of diseases and organisms as well, for example, diseases in ICD-10 are based mostly on body organs or symptoms, but healthcare providers and patients need the pathophysiological or phenotypical information about the disease in order to understand the disease's origin and pathology (NAS, 2011) in order to make more precise treatment decisions.
- The gaps are also due to the focuses or application practices of the KOSs. Classification and thesauri are used to categorize, index, and retrieve information and knowledge contained in research publications, that is, scholarly output from studying the organisms, diseases, diagnosis, and treatments. Scientific taxonomies aim at organizing the knowledge about organisms through applying scientific taxonomy and nomenclature for the purposes of identifying, naming, and classifying them.
- Traditional KOSs are structured and purposed for categorizing and indexing research publications, therefore not equipped with the ability and mechanisms for integrating biomedical research data with taxonomies to form the Knowledge Network of Disease. For example, each class in IDC-10 has the attributes of appellations such as the name(s) of a disease and notation. Similarly, the NCBI Taxonomic classes represent organisms through a larger number of attributes, that is, each class is described by a term ID, inherited blast name, rank, genetic code, other name, host, and lineage.

These gaps separate the representation of basic research data from that of clinical practice and the connections between the two fields are not readily available without making an extra effort. While the goal is to link basic biomedical research to clinical practice, how to build the links becomes a bottleneck problem in the effort to narrow the gaps. What approaches have been used to establish the links between basic biomedical research and clinical practice? Which form may the new generation of KOS take in linking basic biomedical research and clinical practice? These questions will need to be addressed first before we can tackle the grand challenge of building the next generation of KOS that can make an impact on healthcare decision making in a faster and broader manner.

In the following sections, this paper will first review the approaches that have been used to develop the linkage between basic biomedical research and clinical practice, and then describe a pilot study that explores a content analysis approach to lay down the foundation for further data mining. The main purpose of this pilot study is to uncover the kinds of knowledge nodes and relationships that might lead to the

linkage between basic biomedical research and clinical practice, which in turn may be useful for the development of a framework for the Knowledge Network of Disease. As a pilot study, this study is by no means comprehensive nor systematic, but rather, offers a methodological framework for further study.

2 Relevant Research

Terminologies and vocabularies have played a key role for scientific communities to communicate and understand ideas and concepts in the process of knowledge discovery and problem solving, particularly when such activities involve multiple disciplinary fields. KOSs in biomedical domain typically include taxonomies, thesauri, subject heading lists, and ontologies and have been used to taxonomically identify organisms, diseases, DNA and genomic sequences, and other substances, or to represent the subject content of data and/or information resources. Taxonomies in biomedical sciences, for example, describe organisms, diseases, and DNA and genomic sequences identified from biological and medical research. The NCBI Taxonomy (<http://www.ncbi.nlm.nih.gov/taxonomy>) documents organisms by recording the attributes of ID, inherited blast name, rank, genetic code, other names, type material (if any), and full lineage. Other taxonomies such as SNOMED CT and ICD-10 have more or less similar description attributes, i.e., a code and a name. These taxonomies, as the NAS report summarized, are built “primarily based on symptoms, on microscopic examination of diseased tissues and cells, and on other forms of laboratory and imaging studies and are not designed optimally to incorporate or exploit rapidly emerging molecular data, incidental patient characteristics, or socio-environmental influences on disease” (NAS, 2011, p. 14).

Research has attempted to enhance the ability of taxonomies to incorporate molecular data. Using Alzheimer’s disease (AD) as an example, Krishnan (2015) demonstrated how the incorporation of new findings of genetic sequences changed and defined AD by subtyping based on molecular risks. Improving the link between biology and the patient with their treatment and outcome is considered as a “compelling need” (Krishnan, 2015, p. 85). A molecular classification of human cancer can provide a robust taxonomy based on hierarchical grouping, familial syndromes and epidemiologic features, and animal models of the genetic abnormality-etiopathogenic features (Diaz-Cano, 2015). Although research has proven that incorporating molecular research findings does enhance the robustness of taxonomies, they are far from the grand picture of a Knowledge Network of Disease because issues of how to link molecular research data to diseases remain a huge challenge.

The term “knowledge node” is an important concept in this paper, but its use has been primarily in the knowledge management field to denote organizational units, interest groups, communities of practices, or communities of knowing, which “exhibit some degree of semantic autonomy” (Bonifacio, Bouquet, & Cuel, 2002). In the theory of connectivism, knowledge nodes are a network and connect to one another through various types of relationships (AIDahdouh, Osório, & Caires, 2015). Connectivism as a learning theory recognizes three levels of knowledge nodes at which knowledge networks form: 1) neural level at which neural connections form as new stimuli, input, and experiences shape the physical development of the brain; 2) conceptual level that defines the discipline or field of knowledge; and 3) external level that refers to the technologies significantly impacted the formation of networks (Siemens & Tittenberger, 2009). In a knowledge network, nodes may be connected by graded, directional, or self-join relationships or by a set of connections appearing together as a single whole, i.e.,

pattern (AIDahdouh, Osório, & Caires, 2015). Although the definitions of knowledge nodes and networks are not designated for biomedical knowledge networks, the theory about the nodes and relationships offers some insights into how to define knowledge nodes and their relationships for this pilot study.

Different approaches and techniques have been developed in the library and information community to map between different KOSs for integration and enhanced interoperability. Crosswalks have been accomplished for major KOSs used in libraries, for example, the Library of Congress Classification was mapped with Dewey Decimal Classification. In the Linked Data arena, vocabularies from various sources can be integrated via standard identifiers and ontological relationship types. Such processes, however, is full of entangled complications. Partial matches, one-to-many and many-to-one matches, broader and narrower concept matches between the terms in KOSs are common (Zeng & Chan, 2004). Data quality can also affect the accuracy of data integration. For example, variations of identity in Linked Data can result in serious data integration quality issues. In some cases, two identifiers refer to the same thing and all properties hold true for both identifiers, but the same identifiers would not be able to be reused in a different context, just like the same person has different roles in different contexts (Halpin, Herman, & Hayes, 2010).

The Knowledge Network of Disease envisaged by the Committee aims at integrating molecular data, medical histories, and healthcare outcomes for a large number of individuals. The data cumulated from the ongoing delivery of clinical care to these patients will be mined for the construction of this Knowledge Network. Obviously, structures and methods used for traditional KOSs are no longer enough to meet the challenge in building the Knowledge Network of Disease. The vision of a Knowledge Network of Disease represents a departure from traditional KOS in terms of structure, method, and content. While the methods for enhancing KOS integration and interoperability remain useful, new approaches will have to be developed to assure the robustness and usefulness of the Knowledge Network.

3 Data and Methods

3.1 Data Collection

Precision medicine has received an increasing attention from policy makers, biological and clinical researchers, and healthcare providers because of its potential for targeted, precise treatments for diseases. Diseases affecting large populations such as breast cancer and diabetes have received more attention than others. The fact that large numbers of oncogenes have identified to be associated with these prevailing diseases is evidence of active precision medicine research on these diseases. It would make sense for this pilot study to choose research literature on precision medicine for these prevailing diseases to start the ground work.

As stated earlier, this pilot study uses a content analysis approach to discover the types of knowledge nodes and relationships between these nodes from precision medicine literature. A sample of papers was chosen in the area of precision medicine specifically related to two diseases – breast cancer and diabetes – and to one disciplinary field – oncology. Since the goal is to identify what types of knowledge nodes and relationships exist, we decided to use the “purposeful sampling” approach, which is to yield “insights and in-depth understanding rather than empirical generalizations” (Patton, 2002, p. 230). We formulated a few simple queries to search PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) on January 28, 2016, each of which returned

varying numbers of hits when filtered by criteria “Article types: Review”; “Text availability: Full text”; and “Publication dates: 10 years”. The queries and number of hits are as follows:

- precision medicine oncology: 1,377 hits
- precision medicine breast cancer: 202 hits
- precision medicine diabetes: 125 hits

Based on the number of hits for each query, we randomly selected approximately 10% of the papers from each set. This resulted 4 articles for breast cancer, 5 for diabetes, and 11 for oncology, totaling 30 articles. We read through each of the 30 articles to look for potential knowledge nodes and relationships. In purposeful sampling, the researcher will keep sampling the population until the themes or concepts appear to be saturated. It should be pointed out that, although the small numbers of articles selected for each group may not be sufficient to claim a saturation of knowledge nodes and relationships we tried to identify, we did notice repeating knowledge node types and relationships after two or three articles. A knowledge node map for each of the articles was drawn to allow visual comparison between the sampled articles in the same group. Potential knowledge nodes were selected based on the following criteria:

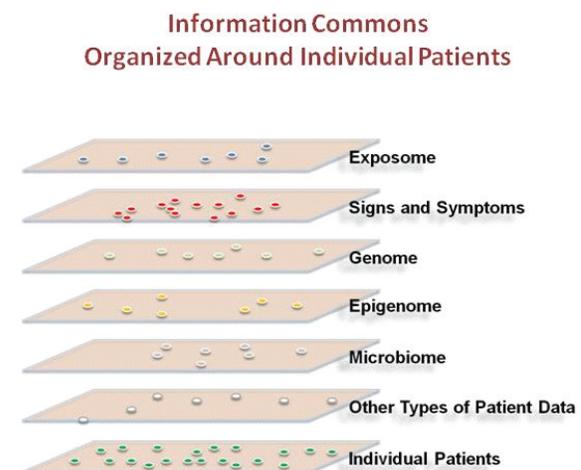
- Molecular entities such as genes, proteins, genomes, etc.
- Disease names
- Names or terms related to treatments/therapies
- Methods, techniques, and types of decisions related to diagnosis
- Data sources used by the publication
- Types of relationships between potential knowledge nodes

3.2 Methods

The NAS report (2011) lays out the complex layers that include molecular data, symptoms, and environmental factors, all of which overlay on the patient layer through complex relationships (Figure 1). While these layers delineate from the patient perspective what data/information should be included in the knowledge network, how each layer interacts with one another will be a grand challenge for making the vision a reality.

As an exploratory pilot study, we tried to incorporate the vision of NAS into our content analysis. For each article, we drew a knowledge map to sort out the relationships between the nodes identified. Figure 2 is one of the maps we drew based on the content of one article. The sample map of knowledge nodes from Figure 2 shows that these knowledge nodes (boxes) may represent a concept at different levels in terms of semantic or hierarchical classes of those concepts, and the relationships between the nodes are often entangled across different levels and categories. For instance, the oncogene EGFR has

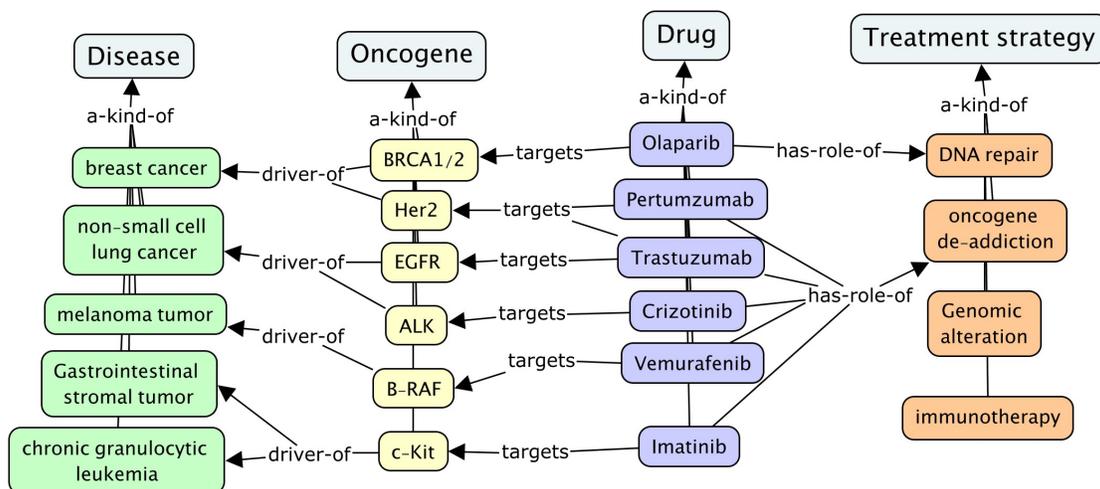
Figure 1. Layers of knowledge nodes
(Source: NAS, 2011, p. 17)



been discovered to associate with a number of different cancers, though only lung cancer is displayed as associated with this oncogene on the map.

It should be pointed out that the purpose of drawing a knowledge map for the sample papers is not to provide an exhaustive listing or map of knowledge nodes and their relationships, but rather, to identify as many types of knowledge nodes and relationships as we can. In fact, not all possible relationships are shown in Figure 2, e.g., genomic alteration and immunotherapy also have relationships with some of the oncogenes, and treatment strategies may be combined to treat cancers at different stages and/or battle secondary resistance to therapies. The types of knowledge nodes and relationships identified from content analysis were collected as the ground work for developing a model or framework that might contribute to the development of the Knowledge Network of Disease, or the New Taxonomy.

Figure 2. A sample map of knowledge nodes and relationships from a research paper (based on PubMed ID 25441102)



4 Results

4.1 Knowledge nodes

A knowledge node in the context of this pilot study is defined as a single concept/name or a set of concepts/names that has at least one connection with another concept/name or set of concepts/names. For example, breast cancer as a kind of disease is a single concept, BRCA1 as a kind of oncogene is a single concept as well as the name of a gene, while mutations in BRCA1 gene can be a set of concepts/names because more than 1,800 mutations in BRCA1 have been identified so far to be associated with increased risk of breast cancer and the ways these mutated genes lead to increased risk of breast cancer are different (NLM, 2016). Based on the pre-determined criteria for selecting potential knowledge nodes and the examination of the sample publication contents, the attributes of knowledge nodes can be grouped by structural levels, disciplinary fields, disease biomarkers, and nodes that blend clinical practice and basic research.

4.1.1 Structural levels

Knowledge nodes are the organics of a knowledge network. In the sample publications, we observed that a large number of knowledge nodes appeared simply as a name for some very specific thing – gene, protein, disease, testing method, and drug, or a term

for a combination of things – e.g., gene damage repairing or alteration, or a term for even a larger scope of combinations, e.g., EGFR mutations in lung cancer, drug resistance through EGFR mutations. Mechanically, the simple names for specific things may be considered as the atom level knowledge nodes. These nodes can exist alone meaningfully and at the same time may be joined together with other atomic nodes to become a concept level node. When several concept nodes form a broader concept or area, the resulting concept may be deemed as a cluster node. Table 1 presents some sample nodes at atomic, conceptual, and cluster levels that were selected from the sample publications:

Table 1. Examples of knowledge nodes derived from the sample publications

Category	Atomic level (name of things)	Concept level	Cluster level
Gene	Her2, BRCA1, BRCA2, EGFR	Oncogenes	EGFR mutations in lung cancer
Disease	Non-squamous carcinoma, squamous cell carcinoma	Non-small cell lung cancer	Lung cancer
Drug	Pertumzumab, Lmatinib, Crizotinib	Tyrosine inhibitor	kinase Oncogene de-addiction

4.1.2 Disciplinary fields

Different types of knowledge nodes may be aggregated broadly by disciplinary field, e.g., genetics, pathology, pathophysiology, oncology, and virology, or more specifically by a domain. The scope of a domain is often difficult to determine, for instance, drug resistance to oncogene de-addiction is a research area that involves types of resistance, cancer stages, genomic patterns, level of intra-tumor heterogeneity, and so on. In this sense, it is similar to cluster-level knowledge nodes.

4.1.3 Disease biomarkers

Unlike ICD in which diseases are named primarily based on symptoms or body organs, the disease names in the sample publications we examined appeared to be closely tied to biomarkers, i.e., genes and proteins. Disease biomarkers are the obvious knowledge nodes. We observed that a disease name often appeared together with genes as disease-biomarker pairs. Examples include chronic myeloid leukemia (CML) with mutated gene BCR-ABL, breast cancer with positive estrogen receptor (ER), BRCA1/2, and Her2, and non-small cell lung cancer with mutations in multiple genes such as epidermal growth factor receptor (EGFR), excision repair-cross complementation group (ERCC), and ribonucleotide reductase (RRM) (Kalia, 2015). The disease-biomarker pairs can be characterized as concept level knowledge nodes, in which the disease node may be associated with multiple biomarker nodes and the same biomarker node may be associated with multiple diseases.

4.1.4 Nodes that blend clinical and basic research

Precision medicine has a strong focus on using information about an individual's genes, proteins, and environment to prevent, diagnose and treat disease (Kalia, 2015). There is a large number of knowledge nodes that can be seen as a blend of clinical practice and basic research. Below are some of the examples from the sample publications we examined:

- clinically actionable mutations
- phenotype of breast cancer
- resistance to endocrine therapy
- biomarkers predicting response to therapy
- genomic drivers of cancer
- predictive and prognostic biomarkers
- intratumor heterogeneity
- molecular classification of tumors

These nodes are usually compound concepts and represent an area or topic of research and/or clinical practice in precision medicine.

4.1.5 Comparing subject terms with knowledge nodes and relationships

Knowledge nodes identified in this pilot study have a different flavor compared to the index terms assigned to the publications. To compare the differences between the knowledge nodes and subject terms from the Medical Subject Headings (MeSH), we collected the terms assigned to the same sample article used in Figure 2 as the example. Table 3 shows that subject term representation is mainly at the concept level, coarser in terms of granularity. Knowledge nodes in Figure 2 are picked directly from the article without “normalization” of terms, thus more straightforward, precise, and likely more closely adhere to the language familiar by researchers and clinicians.

Table 3. Indexing terms for the same sample article as used in Figure 2 (PubMed ID 25441102)

MeSH terms	Substances
Carcinoma, Lung/genetics	Non-Small-Cell EGFR protein, human
Genomic Instability	ERBB2 protein, human
Humans	Proto-Oncogene Proteins c-kit Receptor, Epidermal Growth Factor
Immunotherapy/methods	Receptor, ErbB-2
Lung Neoplasms/genetics	BRAF protein, human
Melanoma/genetics	Proto-Oncogene Proteins B-raf
Molecular Targeted Therapy	
Neoplasms/genetics*	
Neoplasms/therapy*	
Oncogenes	
Precision Medicine/economics	
Precision Medicine/methods*	
Proto-Oncogene Proteins B-raf/genetics	
Proto-Oncogene Proteins c-kit/genetics	
Receptor, Epidermal Growth Factor/genetics	
Receptor, ErbB-2/genetics	
Receptor, ErbB-2/metabolism	
Software	

4.2 Relationships between knowledge nodes

Knowledge nodes are connected through different types of relationships. Typical hierarchical and/or tree relationships such as is-a, a-kind-of, and part-of can be observed from many traditional KOSs, which are also common between the knowledge nodes in the sample publications. Relationship types that are uncommon in traditional KOSs are listed in Table 2. Technically, the relationship types in Table 2 are the results from ontological modeling of a domain.

Table 2. Major relationships types and patterns between knowledge nodes observed in the sample publications

Relationship	Pattern	Example
has-biomarker	<i>Disease</i> has-biomarker <i>Gene</i>	<i>chronic myeloid leukemia</i> has-biomarker <i>BCR-ABL</i> <i>non-small cell lung cancer</i> has-biomarker <i>EGFR</i>
is-driver-of	<i>Gene</i> is-driver-of <i>Disease</i>	<i>Her2</i> is-driver-of <i>breast cancer</i> <i>c-Kit</i> is-driver-of <i>chronic granulocytic leukemia</i>
targets	<i>Drug</i> targets <i>Gene</i>	<i>Crizotinib</i> targets <i>ALK</i> <i>Olaparib</i> targets <i>BRCA1/2</i>
has-role-of	<i>Drug</i> has-role-of <i>Treatment</i>	<i>Crizotinib</i> has-role-of <i>oncogene de-addiction</i> <i>Olaparib</i> has-role-of <i>DNA repair</i>

It should be pointed out that the relationship types in Table 1 are not intended to be exhaustive, but rather, it is to show what most frequently occurred in the sample publications we examined. In the patterns “A has-relationship-type-with B”, both A and B may be substituted with different names or concepts, for example, a treatment may target to a disease, or a gene is-driver-of mutations.

4.3 The knowledge network

From our empirical observation of the sample publications, we noticed that the same knowledge nodes may be marked with different labels – structure, discipline, disease, gene or biomarker, and treatment – and such labeling categorizes them to form different groups or layers in the knowledge network. If we treat these layers as dimensions, then the knowledge nodes are scattered in this multi-dimensional space and connected through different types of relationships. Under this premise, the knowledge nodes in one dimension may be considered as a vector. The sum of all vectors for all dimensions can be computed by using network science algorithms to obtain the knowledge network with desired scale and focuses. As knowledge network is formed by knowledge nodes connected to each other, how these nodes are connected (e.g., strength, distance, first or second order connection) is both a computational problem that requires data science methods to explore and a fundamental problem for KOS research to seek new solutions in building the Knowledge Network of Disease.

5 Discussion

This paper reports a pilot study of a sample of precision medicine publications to uncover and identify the types of knowledge nodes and relationships that may help

generalize a framework or model for building a Knowledge Network of Disease. The different types of knowledge nodes discussed in the result section are not new ideas but inherit some of the theories and methods from KOS and text categorization, an established research field that heavily employs natural language processing, machine learning, and other computational algorithms (Khan et al., 2010; Sebastiani, 2002). What makes the idea of knowledge nodes and relationships used in this pilot study different from traditional KOSs or text categorization lies in the perspective from which knowledge nodes and relationships are defined.

Traditionally, KOSs perform different functions: subject headings such as Medical Subject Headings (MeSH) represent the intellectual content of publications mainly at concept level in the form of discrete subject terms, taxonomies classify and name things in a systematic way, and classification schemes organize knowledge by using coded classes. Each of these KOSs is a world of its own and rarely interact or integrate with one another. On one hand, while these KOSs will remain useful and continue to evolve, their limitations as described in earlier sections are making it difficult for precision medicine to fully utilize the fruit of new molecular biology and other data for more precise healthcare decision making. On the other hand, KOSs' rich semantics can provide "gold standard" for automatically identifying knowledge nodes and relationships among the nodes. In this sense, the knowledge node identification resembles text categorization, except that the nodes will be used to construct the knowledge network.

Biomedical research data are scattered in numerous databases or repositories. Just within the National Center for Biological Information (NCBI), there are 50+ databases. Each of these databases have different data structures and metadata descriptions. Some of them have links to publications. To build the Knowledge Network of Disease, simply linking biomedical datasets to publications is not enough, because such links cannot present a knowledge node and the other nodes and relationships surrounding this node to fully take the opportunity offered by basic research to healthcare decision making. Finer representation of molecular biology and medicine data and knowledge will be a necessary step toward the Knowledge Network of Disease. What will a finer representation look like? This is a question requiring thinking beyond traditional KOSs and the help of big data and metadata analytics.

6 Future study

Although the results from this ongoing pilot study are preliminary and limited in scope, the approach used offers a different way to re-examine KOSs and their merit and limitations as well as the knowledge nodes and their relationships. The findings in this study also need subject experts to validate through further studies. Many questions remain to be explored in building the Knowledge Network of Disease. The key task is to conduct a larger scale study of knowledge nodes and relationships, both manually and automatically, to allow more precise classification and categorization of them. Such a larger scale of study needs to be done to both data (metadata describing the data) and publications and in collaboration with subject experts and data scientists. Networks constructed based on these knowledge nodes and relationships will hopefully help the development of the Knowledge Network of Disease.

References

- AlDahdouh, A. A., Osório, A. J., & Caires, S. (2015). Understanding knowledge network, leaning and connectivism. *International Journal of Instructional Technology and Distance Learning*, 12(10): 3-21. http://www.itdl.org/Journal/Oct_15/Oct15.pdf
- Bonifacio, M., Bouquet, P., & Cuel, R. (2002). Knowledge nodes: the building blocks of a distributed approach to knowledge management. *Journal of Universal Computer Science*, 8(6) (2002): 652-661. http://www.jucs.org/jucs_8_6/knowledge_nodes_the_building/Bonifacio_M.pdf
- Schumann, C.-A. & Tittmann, C. (2013). Correlation between individual knowledge and organizational learning process. In: *Proceedings of the European Conference on Knowledge Management;2013, Vol. 2*, p. 600
- Christiansson, P. (1998). Artificial intelligence in structural engineering. Information technology for design, collaboration, maintenance, and monitoring. In: *Lecture Notes in Artificial Intelligence 1454*. Ian Smith (Ed.), pp. 48-59. Berlin: Springer-Verlag. http://www.it.civil.aau.dk/it/reports/ascona_98/ascona98.html
- Diaz-Cano, S. J. (2015). Pathological bases for a robust application of cancer molecular classification. *International Journal of Molecular Science*, 16(4): 8655-8675. doi: 10.3390/ijms16048655
- Halpin, H., Herman, I., & Hayes, P. J. (2010). When owl:sameAs isn't the same: An analysis of identity links on the Semantic Web. Paper presented at W3C Workshop—RDF Next Steps, Stanford University, Palo Alto, CA, USA. June 26–27, 2010. <http://www.w3.org/2009/12/rdf-ws/papers/ws21>
- Hawgood, S., Hook-Barnard, I.G., O'Brien, T.C., & Yamamoto, K.R. (2015). Precision medicine: Beyond the inflection point. *Science Translation Medicine*, 7, 300ps17.
- Kalia, M. (2015). Biomarkers for personalized oncology: Recent advances and future challenges. *Metabolism*, 64: S16-S21.
- Khan, A., Baharudin, B., Lee, L. H., & Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology*, 1(1): 4-20.
- Krishnan, R. R. (2015). A knowledge network for a dynamic taxonomy of psychiatric disease. *Dialogues in Clinical Neuroscience*, 17(1): 79-87.
- NAS. (2011). *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, D.C.: The National Academies Press.
- NLM. (2016). Genetics Home Reference: Your Guide to Understanding Genetics Conditions: BRCA1 Gene. Website. <https://ghr.nlm.nih.gov/gene/BRCA1>.
- Patton, M. Q. (2002). *Qualitative Evaluation and Research Methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computer Survey*, 34(1): 1-47.
- Simmens, G. & Tittenberger, P. (2009). Handbook of emerging technologies for learning. <http://elearnspace.org/Articles/HETL.pdf>
- Zeng, M. L. & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55(5): 377-395. DOI: 10.1002/asi.10387