

Knowledge Representation and Management: Notable Contributions in 2021

Licong Cui¹, Ferdinand Dhombres^{2,3}, Jean Charlet^{2,4}, Section Editors for the IMIA Yearbook Section on Knowledge Representation and Management

¹ School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA

² Sorbonne Université, INSERM, Univ Sorbonne Paris Nord, LIMICS, Paris, France

³ Sorbonne Université, Service de Médecine Foetale, DMU Origyne, AP-HP, Hôpital Armand Trousseau, Paris, France

⁴ AP-HP, DRCl, Paris, France

Summary

Objectives: To select, present, and summarize the best papers in the field of Knowledge Representation and Management (KRM) published in 2021.

Methods: Following the International Medical Informatics Association (IMIA) Yearbook guidelines, a comprehensive and standardized review of the biomedical informatics literature was performed to select the best KRM papers published in 2021, based on PubMed queries.

Results: A total of 1,231 publications were retrieved from PubMed. We nominated 15 candidate best papers, and four of them were finally selected as the best papers in the KRM section. The topics covered by these papers include knowledge graph, ontology development, ontology alignment, and the International Classification of Diseases.

Conclusions: In the KRM best paper selection for 2021, the candidate best papers covered a wider spectrum of topics compared to the last year's significant focus on ontology curation. In particular, ontology development for specific domains (e.g., Alzheimer's disease, infectious diseases, bioethics) has received the most attention.

Keywords

Knowledge representation and management; ontology; KnowledgeGraph; International Medical Informatics Association

Yearb Med Inform 2022;236-42

<http://dx.doi.org/10.1055/s-0042-1742523>

1 Introduction

Knowledge Representation and Management (KRM) in medicine focuses on the development and application of resources and methods to be used in other medical informatics domains [1-5]. The year 2021 has witnessed a large amount of publications related to KRM in medicine. In this synopsis, we present the selection process of the best papers published in the KRM field for the 2022 International Medical Informatics Association (IMIA) Yearbook, and summarize the nominated candidate best papers according to specific KRM research areas.

2 Paper Selection Method

We conducted the selection of KRM papers published during the year 2021 by querying PubMed/MEDLINE. Our query set includes Medical Subject Headings (MeSH) descriptors related to KRM in the context of medical informatics with a restriction to international peer-reviewed journals, including conference proceedings indexed in PubMed. Only original research articles published from 01/01/2021 to 12/31/2021 were considered; and the following publication types were excluded: reviews, editorials, comments, case reports, and letters to the editors. We reused last year's query set [5] with "semantics" added this year. Note that we did not search the Web of Science database due to its limited number of KRM publications observed in 2018 and 2019 [5].

We followed the generic method commonly used in all sections of the IMIA Yearbook to select the best papers. The selection process consisted of three steps. Firstly, the section editors reviewed the title, abstract, and publication types of all the retrieved articles to select a short list of 15 candidate best papers. Secondly, each candidate paper was peer-reviewed by at least four expert reviewers, consisting of two section editors, one editor in chief, and one or two external reviewers. The peer review was performed according to the IMIA Yearbook's standard evaluation criteria, including significance, quality of scientific and/or technical content, originality and innovativeness, coverage of related literature, and organisation and clarity of presentation. Thirdly, the final decision of the best papers was reached during a meeting of the whole editorial board, based on the peer reviews and the report of the section editors.

3 Results

3.1 Best Paper Selection for 2021

We retrieved a total of 1,231 KRM related publications in 2021 from PubMed, which is slightly more than last year's (1,175). The section editors' initial screening based on the title and abstract resulted in 326 papers. This set of papers was further reviewed jointly by the section editors to select a consensus list of 15 candidate best papers [6-20]. The four top-ranked papers according

to expert reviewers' evaluation were selected as the best papers in the field of KRM published in 2021 (see Table 1).

The first paper is a contribution by Vogt [6], who performed a comprehensive comparison of class-based TBox representation (i.e., ontology) and instance-based ABox representation (i.e., knowledge graph) to document and manage empirical data and metadata for anatomical research, in compliance with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. It was concluded that the ABox approach seems to be in general superior to the TBox approach for representing and managing empirical data.

In the second article, Keet and Grütter [7] proposed a framework to systematically handle modelling conflicts via meaning negotiation and conflict resolution in the ontology development and (re)use processes. They introduced a preliminary library of conflicts that may emerge during ontology authoring, and the conflict set data structure that stores the minimum necessary data about such conflicts. Then they proposed resolution strategies and general principles for resolving different types of conflicts.

The third paper is a contribution by Wang *et al.* [8], who investigated and compared the effectiveness of different combination strategies for matching biomedical ontologies. They represented multiple matchers in four dimensions: terminology, structure, external knowledge, and representation learning. Their experimental results showed that the combination of all

four dimension-based matchers achieved the best performance for matching the Adult Mouse Anatomy to the National Cancer Institute (NCI) Thesaurus; while the combination of terminology, structure, and external knowledge-based matchers achieved the best performance for matching the Foundational Model of Anatomy (FMA) to the NCI Thesaurus and for matching the FMA to the SNOMED CT.

In the fourth article, Harrison *et al.*, [9] provided an overview of the recently completed 11th revision of the International Classification of Diseases (ICD-11), highlighting the main innovations and important features of ICD-11 in comparison with earlier versions. A new information framework has been introduced in the ICD-11, consisting of a semantic knowledge base (the Foundation), classifications derived from the Foundation, and a common biomedical ontology linked to the Foundation.

Additional content summaries of the four best papers can be found in the Appendix.

Figure 1 shows a tag cloud of the 15 candidate best papers based on a curated list of keywords for each paper. Among all the 15 candidate best papers in KRM for 2021, we observed four main research areas: knowledge graph, ontology development, ontology alignment, and International Classification of Diseases, covering a wider range of topics in contrast with the KRM selection for the year 2020, which observed a significant focus on methods and tools for ontology curation.

3.2 Knowledge Graph

In addition to the best paper from Vogt [6], which recommends using a knowledge graph rather than an ontology to represent empirical data in anatomy, there are three other candidate papers about building knowledge graphs for disparate domains [10-12].

In the candidate paper from Delmas *et al.*, [10], the authors developed an openly accessible knowledge graph (named FORUM) by extracting the associations between chemical entities and biomedical concepts in public databases (PubChem, MeSH, ChEBI, ChemOnt and MetaNetX) and biomedical literature (PubMed) for metabolomics research. This is a useful resource supporting metabolomics analysis, result interpretation, and hypothesis generation.

Deng *et al.*, [11] proposed PhenoSSU, a fine-grained semantic information model (an "entity-attribute-value" model) for representing phenotype knowledge in clinical guidelines for infectious diseases. They have constructed PhenoSSU-based knowledge graphs for 193 infectious diseases with 4,020 PhenoSSU instances. The PhenoSSU model outperformed the clinical element model and HL7 fast healthcare interoperability resource (FHIR) model when comparing their expressive power to capture the full semantics underlying the natural language phenotype descriptions listed in clinical guidelines.

Huang *et al.*, [12], reported their work on constructing knowledge graphs of Kawasaki disease by integrating various knowledge sources, including clinical trials, PubMed papers, medical guideline, drug bank, drug side effect, and SNOMED-CT. The authors also showed several use cases how these knowledge graphs can be used for supporting efficient semantic search in the study of Kawasaki Disease. However, more systematic and quantitative evaluation of this efficiency is still needed by follow-up research.

3.3 Ontology Development

Ontology development remained an active KRM research area in the year 2021. Among the 15 candidate best papers, eight of them are with regard to development of ontologies. While the best paper from Keet and Grütter [7] focused on studying methodologies to

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2022 in the section 'Knowledge Representation and Management'. The articles are listed in alphabetical order of the first author's surname.

Section
Knowledge Representation and Management
<ul style="list-style-type: none"> ▪ Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. <i>BMC Med Inform Decis Mak</i> 2021;21(Suppl 6):206. ▪ Keet CM, Grütter R. Toward a systematic conflict resolution framework for ontologies. <i>J Biomed Semantics</i> 2021;12:15. ▪ Vogt L. FAIR data representation in times of eScience: a comparison of instance-based and class-based semantic representations of empirical data using phenotype descriptions as example. <i>J Biomed Semantics</i> 2021;12:20. ▪ Wang P, Hu Y, Bai S, Zou S. Matching Biomedical Ontologies: Construction of Matching Clues and Systematic Evaluation of Different Combinations of Matchers. <i>JMIR Med Inform</i> 2021;9(8):e28212.

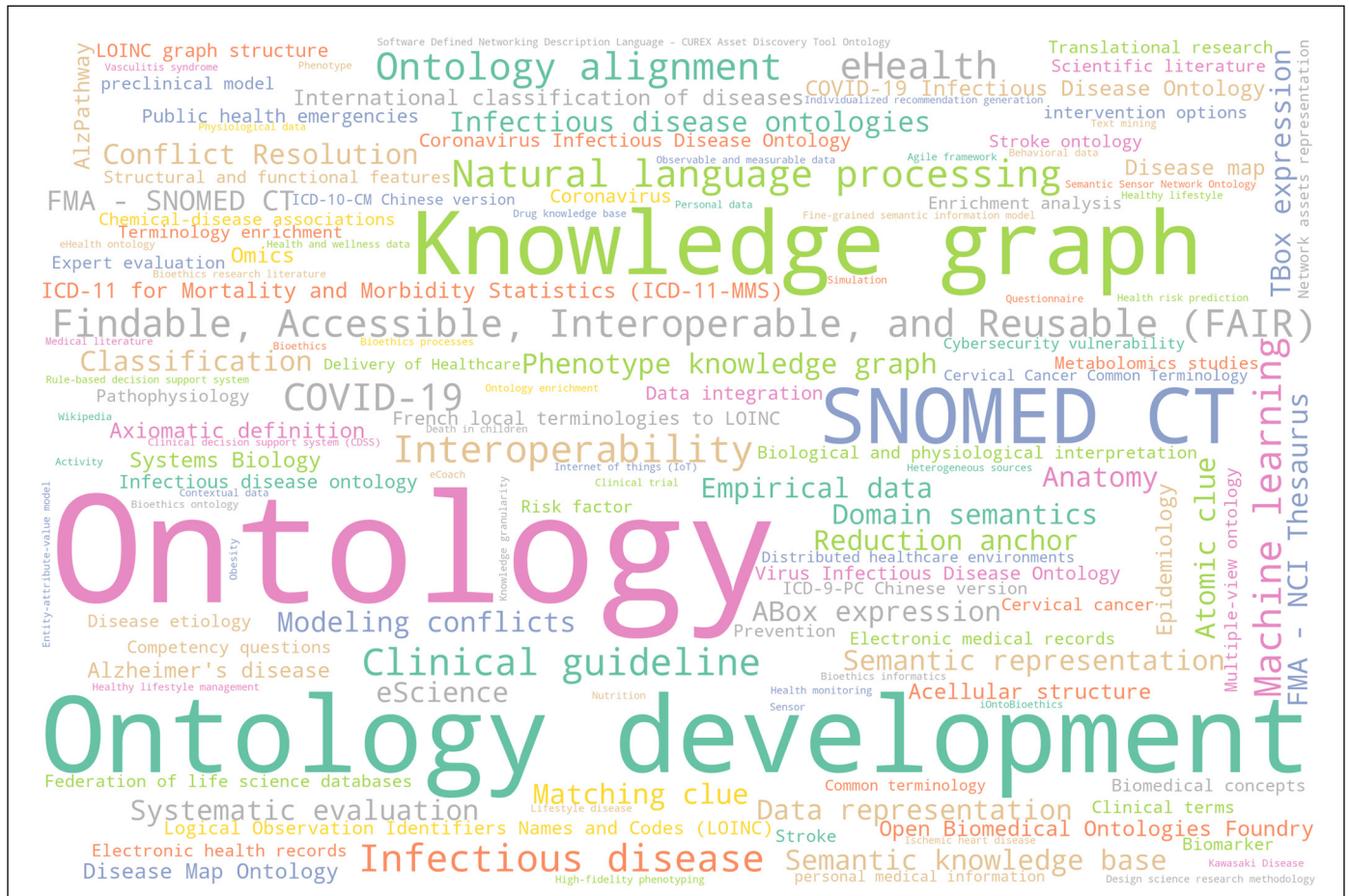


Fig. 1 Tag cloud of curated keywords for the 15 candidate best papers.

resolve modelling conflicts when reusing multiple ontologies, the other seven candidate papers contributed to the construction of domain-specific ontologies or terminologies, including Alzheimer's disease (AD), infectious diseases, stroke, cervical cancer, computer network assets in healthcare settings, bioethics, and lifestyle diseases.

Henry *et al.*, [13] proposed an ontological upper model called Disease Map Ontology (DMO) to convert disease maps to formal ontologies. Such conversion was needed due to the limited expressiveness of systems medicine disease maps, which may lead to errors and misinterpretations. The authors illustrated DMO's utility for AD by converting AlzPathway (a disease map developed for AD) to Alzheimer DMO (a formal ontology). The resulting formal ontologies have been

made freely available in BioPortal. The proposed approach is generally applicable to transform other disease maps to ontologies.

The paper from Babcock *et al.*, [14], describes in detail the recent evolutions made on the Infectious Disease Ontology (IDO) to support data integration and analysis for COVID-19 and more general viral infectious diseases. Based on the existing IDO Core, the authors have developed three new extensions: the Virus Infectious Disease Ontology (VIDO), the Coronavirus Infectious Disease Ontology (CIDO), and the COVID-19 Infectious Disease Ontology (IDO-COVID-19), which have been shared publicly on GitHub, Ontobee and BioPortal. These new extensions have been used for annotating COVID-19 clinical trials, epidemiological, and pathogenesis data in the U.S.

National Library of Medicine COVID-19 corpus. Through the building of the extensions, the authors illustrated that IDO Core provides a simple recipe for building new pathogen-specific ontologies, which can be helpful for rapidly responding to the research needs for future threat of novel pathogens.

Habibi-Koolae *et al.*, [15], developed the Stroke Ontology (STO) to represent knowledge in the domain of brain stroke from multiple perspectives, including risk factors, prevention, disease etiology, pathophysiology, biomarkers, preclinical models, and intervention options. It overcomes the limitation of other existing stroke-related ontologies and classification systems, which have focused on a single clinical view. STO is a useful resource for the stroke research community. It has been made freely available on BioPortal.

In the candidate paper from Hong *et al.*, [16], the authors have developed the Cervical Cancer Common Terminology (CCCT) to facilitate data analysis and exchange for cervical cancer clinical research in China. The key domain concepts were identified from clinical guidelines and medical books by manual review of clinical experts. Term enrichment was further performed using both machine learning and rule-based natural language processing techniques to extract terms from clinical notes.

Santamaría *et al.*, [17] developed an ontology for representing computer network assets and features in healthcare environments, called the Software Defined Networking Description Language - CUREX Asset Discovery Tool Ontology (SDNDL-CAO). The ontology was designed to model data from distributed healthcare environments regarding devices and networks' topologies within the context of detecting potential cybersecurity vulnerabilities. However, the applicability and validation of the ontology in the real healthcare setting need further investigation.

In the candidate paper from Odeh *et al.*, [18], utilizing over 26,000 articles related to bioethics processes indexed by Scopus, the authors developed the iOntoBioethics ontology through two ways: manual construction, and automatic generation using text mining and machine learning. Domain expert validation found that the two approaches complemented each other, with the automatic approach generating concepts at a higher level of abstraction while the manual approach providing more detailed and specific concepts at a lower level of abstraction. Unification of the two approaches' outcome produced the final iOntoBioethics ontology. Additional extensions of this work resulted in the iOntoBioethics COVID-19 pandemic ontology. These are valuable resources for the bioethics domain.

Chatterjee *et al.*, [19] proposed an eHealth ontology to integrate and annotate personal, physiological, behavioral, and contextual data generated from heterogeneous sources such as internet of things (IoT) sensors, questionnaires, and interviews, in the context of individualized health risk prediction for lifestyle diseases like obesity. The eHealth ontology leveraged relevant concepts in the SNOMED-CT and the Semantic Sensor Network Ontology. The authors targeted obe-

sity as a study case and used artificial data simulated in the health monitoring setting. This is a proof-of-concept study supporting healthy lifestyle management.

3.4 Ontology Matching

Various matching or alignment strategies for biomedical ontologies and terminologies have been investigated during the last decade, as illustrated by two of the 15 candidate best papers selected for the year 2021.

In the best paper from Wang *et al.*, [8], the authors performed a systematic empirical study to compare different combinations of matching strategies, which has shed light on the varying effects of matching dimensions used for conducting disparate matching tasks.

The candidate paper from Nikiema *et al.*, [20] describes the construction of a graph representation of the Logical Observation Identifiers Names and Codes (LOINC) incorporating all its French language variants. The built LOINC graph structure is based on the labels of LOINC. It not only can be used to facilitate the alignment of French local terminologies to LOINC, but also lays a foundation for subsequent related studies, such as quality assessment of different (French) translations of LOINC.

3.5 The International Classification of Diseases

We categorized the International Classification of Diseases as a dedicated research area given its long history, large-category classifications, and wide usage. The best paper from Harrison *et al.*, [9] reports the detailed content changes and improvements of the ICD-11 in comparison to the ICD-10, which was developed about 30 years ago. The improvements include a new content model allowing easy incorporation of new entities, supporting the combination of codes to form clusters, enhanced interoperability in digital health information environments, and a web-based coding tool for users to find and select categories. This is a rather informative paper summarizing substantial design and structure changes of the ICD-11 to overcome shortcomings of the ICD-10.

4 Conclusions

The 15 candidate best papers selected for the year 2021 illustrated a wider spectrum of KRM research areas: knowledge graph, ontology development, ontology alignment, and International Classification of Diseases, with each area featuring a best paper. Nearly half of the candidate papers were devoted to the development of domain-specific ontologies, including Alzheimer's disease, infectious diseases, stroke, cervical cancer, lifestyle diseases such as obesity, computer network assets in healthcare settings, and bioethics.

Acknowledgements

We would like to acknowledge the support of Fleur Mougin, Adrien Ugon, Lina Soualmia, Kate Fultz Hollis, and the whole IMIA Yearbook editorial team as well as the reviewers in the selection process of the 2021 KRM best papers.

References

1. Dhombres F, Charlet J. Knowledge Representation and Management, It's Time to Integrate! *Yearb Med Inform* 2017 Aug;26(1):148-51.
2. Dhombres F, Charlet J; Section Editors for the IMIA Yearbook Section on Knowledge Representation and Management. As Ontologies Reach Maturity, Artificial Intelligence Starts Being Fully Efficient: Findings from the Section on Knowledge Representation and Management for the Yearbook 2018. *Yearb Med Inform* 2018 Aug;27(1):140-5.
3. Dhombres F, Charlet J; Section Editors for the IMIA Yearbook Section on Knowledge Representation and Management. Formal Medical Knowledge Representation Supports Deep Learning Algorithms, Bioinformatics Pipelines, Genomics Data Analysis, and Big Data Processes. *Yearb Med Inform* 2019 Aug;28(1):152-5.
4. Dhombres F, Charlet J; Section Editors for the IMIA Yearbook Section on Knowledge Representation and Management. Design and Use of Semantic Resources: Findings from the Section on Knowledge Representation and Management of the 2020 International Medical Informatics Association Yearbook. *Yearb Med Inform* 2020 Aug;29(1):163-8.
5. Dhombres F, Charlet J; Section Editors for the IMIA Yearbook Section on Knowledge Representation and Management. Knowledge Representation and Management: Interest in New Solutions for Ontology Curation. *Yearb Med Inform* 2021 Aug;30(1):185-90.
6. Vogt L. FAIR data representation in times of eScience: a comparison of instance-based and

- class-based semantic representations of empirical data using phenotype descriptions as example. *J Biomed Semantics* 2021 Nov 25;12(1):20.
7. Keet CM, Grütter R. Toward a systematic conflict resolution framework for ontologies. *J Biomed Semantics* 2021 Aug 9;12(1):15.
 8. Wang P, Hu Y, Bai S, Zou S. Matching Biomedical Ontologies: Construction of Matching Clues and Systematic Evaluation of Different Combinations of Matchers. *JMIR Med Inform* 2021 Aug 19;9(8):e28212..
 9. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. *BMC Med Inform Decis Mak* 2021 Nov 9;21(Suppl 6):206.
 10. Delmas M, Filangi O, Paulhe N, Vinson F, Duprier C, Garrier W, et al. FORUM: Building a Knowledge Graph from public databases and scientific literature to extract associations between chemicals and diseases. *Bioinformatics* 2021 Sep 3;37(21):3896–904.
 11. Deng L, Chen L, Yang T, Liu M, Li S, Jiang T. Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study. *J Med Internet Res* 2021 Jun 15;23(6):e26892.
 12. Huang Z, Hu Q, Liao M, Miao C, Wang C, Liu G. Knowledge Graphs of Kawasaki Disease. *Health Inf Sci Syst* 2021 Feb 27;9(1):11..
 13. Henry V, Moszer I, Dameron O, Vila Xicota L, Dubois B, Potier MC, et al; INSIGHT-preAD Study Group. Converting disease maps into heavyweight ontologies: general methodology and application to Alzheimer's disease. *Database (Oxford)* 2021 Feb 16;2021:baab004.
 14. Babcock S, Beverley J, Cowell LG, Smith B. The Infectious Disease Ontology in the age of COVID-19. *J Biomed Semantics* 2021 Jul 18;12(1):13.
 15. Habibi-Koolae M, Shahmoradi L, Niakan Kalhori SR, Ghannadan H, Younesi E. STO: Stroke Ontology for Accelerating Translational Stroke Research. *Neurol Ther* 2021 Jun;10(1):321-33.
 16. Hong N, Chang F, Ou Z, Wang Y, Yang Y, Guo Q, et al. Construction of the cervical cancer common terminology for promoting semantic interoperability and utilization of Chinese clinical data. *BMC Med Inform Decis Mak* 2021 Nov 16;21(Suppl 9):309.
 17. Prieto Santamaría L, Fernández Lobón D, Díaz-Honrubia AJ, Ruiz EM, Nifakos S, Rodríguez-González A. Towards the Representation of Network Assets in Health Care Environments Using Ontologies. *Methods Inf Med* 2021 Dec;60(S 02):e89-e102.
 18. Odeh M, Kharbat FF, Yousef R, Odeh Y, Tbaishat D, Hakooz N, et al. iOntoBioethics: A Framework for the Agile Development of Bioethics Ontologies in Pandemics, Applied to COVID-19. *Front Med (Lausanne)* 2021 May 21;8:619978.
 19. Chatterjee A, Prinz A, Gerdes M, Martinez S. An Automatic Ontology-Based Approach to Support Logical Representation of Observable and Measurable Data for Healthy Lifestyle Management: Proof-of-Concept Study. *J Med Internet Res* 2021 Apr 9;23(4):e24656.
 20. Nikiema JN, Mougín F, Jouhet V. Building a Graph Representation of LOINC® to Facilitate its Alignment to French Terminologies. *AMIA Annu Symp Proc* 2021 Jan 25;2020:933-42.

Correspondence to:

Licong Cui
 School of Biomedical Informatics
 The University of Texas Health Science Center at Houston
 7000 Fannin Street
 Houston, TX 77030, USA
 E-mail: licong.cui@uth.tmc.edu