This is a previous version of the article published in Data & Knowledge Engineering. 2021, 135: 101911. https://doi.org/10.1016/j.datak.2021.101911

Conceptual Modeling in the Era of Big Data and Artificial Intelligence: Research Topics and Introduction to the Special Issue

Juan Trujillo[a], Karen C. Davis[b], Xiaoyong Du[c], Ernesto Damiani[1], Veda C. Storey[d]

[a]*Lucentia Research Group, Department of Software and Computing Systems, University of Alicante, Alicante - 03690, Spain*
[b]*Computer Science and Software Engineering Department, Miami University, Oxford, OH, USA 45056*
[c]*School of Information, Renmin University of China, Beijing, China 100872*
[d]*J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia, United States 30302-4015*

Abstract

Since the first version of the Entity-Relationship (ER) model proposed by Peter Chen over forty years ago, both the ER model and conceptual modeling activities have been key success factors for modeling computer-based systems. During the last decade, conceptual modeling has been recognized as an important research topic in academia, as well as a necessity for practitioners. However, there are many research challenges for conceptual modeling in contemporary applications such as Big Data, data-intensive applications, decision support systems, e-health applications, and ontologies. In addition, there remain challenges related to the traditional efforts associated with methodologies, tools, and theory development. Recently, novel research is uniting contributions from both the conceptual modeling area and the Artificial Intelligence discipline in two directions. The first is efforts related to how conceptual modeling can aid in the design of Artificial Intelligence (AI) and Machine Learning (ML) algorithms. The second is how Artificial Intelligence and Machine Learning be applied in model-based solutions, such as model-based engineering, to infer and improve the generated models. For the first time in the history of Conceptual Modeling (ER) conferences, we encouraged the submission of papers based on AI and ML solutions in an attempt to highlight research from both communities. In this paper, we present some of important topics in current research in conceptual modeling. We introduce the selected best papers from the 37th International Conference on Conceptual Modeling (ER'18) held in Xi'an, China and summarize some of the valuable contributions made based on the discussions of these papers. We conclude with suggestions for continued research.

Keywords: conceptual modeling, big data, machine learning, artificial intelligence

1. Introduction

Since the appearance of the Entity-Relationship Model by P.P.S. Chen [11], research and applications of conceptual modeling have continued to expand. Conceptual modeling can be described in general terms as a body of knowledge about abstraction techniques for representing artifacts, and their semantics, that are associated with software. The structure (entities, relationships, attributes), constraints, and states (and transitions or transformations between states) of software artifacts can be expressed by conceptual models. Conceptual models are used by practitioners as a means of documenting software applications and as basis for development or implementation of systems. They can be employed for description, design, quality assurance, and automatic creation of software artifacts. A classification of conceptual modeling research provides an overview of many facets of the field [15]. Researchers are continuously extending conceptual modeling techniques and approaches to address new challenges that arise as a result of advances in computing technologies and the evolving demands of contemporary applications.

The *International Conference on Conceptual Modeling* (formerly Entity Relationship) is the leading conference for the presentation and exchange of ideas and concepts that relate to traditional and emerging issues in conceptual modeling of information systems, attracting both researchers and business practitioners. Work on conceptual modeling has continued to evolve as the ER model has been applied, modified, and extended to research in database management systems, business process management, and management information systems. Advances in conceptual modeling have also occurred in ORM (Object-Relational Mapping), UML (Unified Modeling Language), i-star, and BPMN (Business Process Model and Notation), and other initiatives. Conceptual modeling is continuing to play a vital role in the emerging, new data era where the correct design and development of mobile or sensors analytics, Big Data systems, decision support systems, NoSQL databases, smart cities, and biomedical systems will be crucial.

The main research topics of the ER'13 conference [72] suggested forthcoming novel research topics. The authors highlighted the relevance of conceptual modeling in areas such as ontologies and Big Data. Eight years later, we continue to highlight the current relevance of conceptual modeling in these areas. In addition to our review of recent work, we include the topic of Artificial Intelligence and conceptual modeling, which will be a highly relevant topic for the next decade.

This introduction to the *Data and Knowledge Engineering* Special Issue on ER 2018 first identifies some of the important and emerging areas of research in conceptual modeling as reflected in that meeting. It then provides an overview of each of the papers that appear in this special issue before concluding with a summary and discussion of future aspects of conceptual modeling.

2. Research Topics

Conceptual modeling continues to be relevant in this new era when computing systems are dominated by Big Data and Artificial Intelligence. Within these two main areas, there are many different emerging research topics. Conceptual modeling efforts continue to demonstrate that they play a critical role as is evident from the vast amount of different software systems, algorithms, data processing and analysis architectures being developed.

The main research topics that emerged from the ER 2018 conference [78] included papers related to fundamentals of conceptual modeling, ontologies, semi-structured and spatio-temporal modeling, Big Data modeling approaches, language and models, and conceptual modeling for machine learning. Thus, and after carefully grouping the contributions and making an intersection of the main research topics covered by the accepted papers, we have decided to provide a summary of the state of the art and our own perspectives in three main areas: (i) Big Data and conceptual modeling, (ii) Machine Learning and conceptual modeling, and (iii) ontologies and conceptual modeling.

2.1. Big Data and Conceptual Modeling

One of the main challenges of Big Data, for both researchers and practitioners, is its tremendous landscape based on the vast number of applications, solutions, and architectures. These can be applied based on various criteria, such as data sources, types of analysis to be conducted, or the role of the users. Prior research has recognized the need for conceptual modeling to support Big Data research (e.g. [19], [85], [53]).

There are many definitions for Big Data. Commonly, Big Data is the way to process or analyze volumes of data that from its unstructured nature and size cannot be analyzed in a timely way using traditional Business Intelligence applications. From Business Intelligence, Big Data has been a natural evolution of processing and analyzing data for decision-making purposes when the source of data is characterized as having a large volume and heterogeneity [10]. It is widely accepted that Big Data was initially defined by the 3 V's of Big Data: Volume, Velocity, and Variety. *Volume* refers to the huge volumes of data available for processing, with orders of magnitude of terabytes, petabytes or even exabytes; *Velocity* refers to the dramatic increase of data generated in (near) real time and constraints on the time available for processing them; and *Variety* refers to a significant increase in the number of heterogeneous data sources that must be integrated with differing and less structured data models. More V's have been added to the definition of Big Data, highlighting Veracity [13, 16] and Value [16, 86]. *Veracity* refers to the trustworthiness of the data source and it is highly related to the data quality of the data source, whereas *Value* refers to the added benefit that the new Big Data source can add to the decision-making process. In total, there are 11 V's to define and properly classify Big Data. Prior research has attempted to identify the potential role of conceptual modeling as it relates to big data (e.g. [69]). There is also research on ontology and semantic interoperability to deal with the variety issues that stems from heterogeneous data sources (e.g. [9]) and ontologies and the modeling of value (e.g. [3], [68]).

The keynote entitled *Recent Trends of Big Data Platforms and Applications* by Kyu-Young Whang [17] gave an overview of the current trends for Big Data platforms and applications from the practitioner's point of view. Whang emphasized that researchers face many remaining research challenges that can be directly applied in current real-world scenarios. One challenge is that an aim of Big Data systems is to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

After reviewing the main Big Data sources and scenarios, such as the Internet of Things (IoT), Web 2.0, scientific experiments, mobile data, and healthcare data, Whang provided an overview of the state of the art regarding current Big Data platforms and applications. He presented the Tencent Distributed Database System (TDDS), and the Object-Relational DBMS (Database and Management System) developed at KAIST (Korea Advanced Institute for Science and Technology), showing that these systems are massively parallel search engines with higher functionality than a DB-IR (Database and Information Retrieval) tightly integrated parallel DBMS. He summarized the main research challenges creators encountered as well as the key role of conceptual modeling in properly solving some of the encountered problems. He especially emphasized that emerging applications are realizing Big Data intelligence, thus, decreasing the gap between Big Data and Artificial Intelligence. Thus, conceptual modeling of Big Data and Machine Learning applications will need to work together to realize mutual benefits in the near future [52].

Recently, there have been several methodological proposals [58, 41, 1, 57, 21] providing effective solutions that help practitioners to develop Big Data solutions. Most of these proposals are based on the analysis of the requirements derived from the 5V's or characteristics of Big Data, which helps with the definition of the Big Data pipeline. Requirements modeling approaches [58, 41] are based on the premise that one of the most common causes of the failure of Big Data projects is the lack of clear and measurable requirements. There are other approaches that use requirements modeling in combination with Data Warehousing (DW) techniques traditionally used in Business Intelligence (BI) systems [40, 48, 54]. The applicability of existing DW methodologies in Big Data scenarios is analyzed and compared [21] to provide guidance on how the traditional BI/DW applications should evolve to properly consider Big Data scenarios.

Another methodological proposal is the TOREADOR project [14] (EU Horizon 2020), which provides a methodology to define the requirements of a Big Data system through a declarative language. From this definition, the TOREADOR methodology allows designers to automatically generate the implementation of the architecture by using an ontology for the selection of the technologies. In a related aspect of the TOREADOR project, a methodology based on Model Driven Engineering is proposed for the management of a Big Data pipeline with support for the automation of Big Data Analytics processes [1]. These research efforts demonstrate that conceptual modeling is highly relevant to Big Data.

Other methodological approaches [75, 76] complement prior proposals by paying particular attention to the workflow of the Big Data architectures from a designer's perspective. These methodologies are based on the fact that one of the main problems of Big Data projects is the lack of expertise needed to combine and select the right technologies to build the correct Big Data architecture for a given problem. The latter version of this step-by-step methodology allows the generation of Big Data pipelines based on several requirements derived from source features that are critical for the selection of the most appropriate tools and techniques. Testing of the methodology on real-world scenarios shows that it reduces the required know-how from practitioners.

From our perspective, we consider Business Intelligence (BI) applications as the precursors of Big Data applications, mainly because BI scenarios have long recognized the need to deal with heterogeneous data sources that must be processed, transformed, and then integrated for decision-making purposes. Considering the history of research in Data Warehouse (DW) applications, and acknowledging the continuous increasing complexity of BI applications, the BI research community started proposing several conceptual modeling approaches to facilitate the modeling of these applications. The most relevant conceptual models appeared in the late 90s and are now being extended. Two representative and well-known conceptual models for DWs [48, 28, 77] use different notation and language foundations; however, they model DWs in an abstract way. Thus, as has happened with DWs in the past, there are still many remaining research challenges for the modeling of Big Data scenarios, so the conceptual modeling of Big Data sources will, most likely, remain an important area of area during the next decade.

Finally, another critical issue in Big Data scenarios that has received a lot of attention recently from both the research and practitioner communities, is the visualization of Big Data sources. Not all the visualization techniques are suitable for all the Big Data sources and users. Therefore, it is necessary to analyze different factors before visualizing Big Data sources. Examples of relevant factors include: the type of source, the role of users and their expertise, the goal of the analytic process, and the frequency with which the data can, or must, be analyzed. In this research direction, approaches have been, or continue to be, proposed to automate data visualization from user requirements [18, 62, 44]. The Model Driven Architecture (MDA) approach automates the derivation of the most appropriate visualization from user requirements [27]. Another model-driven approach automatically generates the most suitable visualization from user's requirements in Big Data scenarios by following a Goal-oriented approach based on the i-star notation [43].

Visualization is also crucial in both Big Data and Machine Learning. It is widely accepted that one of the research challenges in Machine Learning (ML) algorithms and techniques is to facilitate the interpretation of their results by non-expert users. Therefore, the visualization of Machine Learning algorithms will be continue to be an important research topic. An example of visualization for both topics, is an approach to automatically detect and visualize bias in data analytics to address bias in the output of machine learning algorithms [42].

2.2. Machine Learning and Conceptual Modeling

In the ER'18 conference, our keynote speaker Ernesto Damiani gave a presentation entitled *Towards Conceptual Models for Machine Learning Computations*. In this section, we summarize the main issues covered by this keynote.

The proliferation of Internet-of-Things (IoT) devices and the distributed nature of data sources has deeply changed the role of metadata design. By supplying information pertaining to the nature of the data items being generated, and the relations between them, today's metadata encodes essential knowledge for data querying as well as for data integration, data quality and governance.

The advent of Machine Learning (ML) is affecting conceptual modeling in two ways: firstly, by substituting or complementing the symbolic inference algorithms traditionally underlying data services with machine learning models; secondly, by blurring the distinction between data instances and schema in ML-supported query and analysis.

For the first aspect, ML is quickly gaining importance in automating data integration tasks. Embedding ML components in data integration processes can support ontology-based data integration techniques based on computing common abstractions, called (*semantic lifting*). Examples include integrating event logs across an organization without a shared activity ontology [39], and automating data cataloging and data characterization (e.g., inferring metadata from data content [2]).

For the second aspect, conceptual modelling needs to support the design of metadata suitable for *data-metadata fusion*; that is, to create a common data space where ML models can be trained. Data-metadata fusion puts crucial conceptual information within the reach of ML models' perception, including the characteristics of the data points, their relationships, trustworthiness, and past usage. Recently, ML methods have been developed to "homogenize" data and metadata creating a unitary data space (technically, a *common manifold*). By sampling in the manifold, we can obtain inputs on which the actual inference operates[1].

An important research topic is how to combine data and metadata so that the manifold formation process automatically focuses on information from the most reliable data. This type of *dynamic fusion* is difficult to perform when feeding ML models to be trained in the traditional way (i.e., train first, and use later) because it requires the data-metadata fusion to take place simultaneously with the use of the ML model.

From this consideration, emerged the idea of using ML models in a generative way, directly producing data in the integrated data-metadata space with a strategy that takes into account of the different levels of reliability of the data, encoded in the corresponding metadata. To use this strategy, the possible data

[1]It is worth remarking that, according to some neurologists [26], the current trend in computer science research toward early integration at the perceptual level of data and metadata (*early fusion*) is profoundly different from the multi-sensory integration strategies of the human brain.

manifolds from which to feed the integrated model are selected by proceeding in an unsupervised way; that is, by trial and error. This way, concepts will help ML models to navigate unfamiliar environments and contextualize new knowledge and experiences.

The ER'18 conference had two sections devoted to the topic of Conceptual Modeling and Machine Learning. The papers presented in these two sections covered the topics of improving knowledge discovery processes or improving the description of techniques used in Machine Learning components or systems [36]. In the area of improving knowledge discovery, the topics included: deep learning techniques for incorporating reviews into recommendation systems [37]; mining rules with constants for knowledge base construction [83]; an approach for extracting structural relationships [51]; real-time event summarization from microblogging data [47]; natural language text classification [35]; and discovering regular expressions for schemas in XML documents [45].

2.3. Ontologies and Conceptual Modeling

Ontology relates to the study of existence and is a branch of philosophy that examines the fundamental nature of being. These is much overlap between ontology and conceptual modeling because both deal with understanding concepts in the real world. Many of the applied ontology applications start by creating a domain ontology. Indeed, Guarino [29] has long recognized that all information systems have ontologies that are not explicit, but embedded in parts of the systems. For conceptual modeling, research on ontologies has become an important, and active area of inquiry over the past fifteen years [46].

There is much ongoing research into the theoretical and philosphical basis of ontology for conceptual modeling ([31], [30], [32]). A large body of work on the Foundational Ontology has been undertaken by Guizzardi and his colleagues (e.g., [34]) as well as work on ontological clarity for conceptual modeling constructs [24]. Other efforts have identified the usefulness of applying ontology to conceptual modeling (e.g., [33], [82], [59], [65], [81], [84]). The Bunge Wand Weber ontology, based on the ideas of the philosopher physicist Bunge [7] has been widely used in a wide variety of applications (e.g., [60], [50]). For domain ontologies, many significant issues are being, or have been researched. Topics include development and representation ([38], [22], [64]); construct analysis ([80], [63], [71]); application and use ([5], [23], [20], [70], [12], [73], [74]); and assessment (e.g., [8]; [56]; [55] as based on semiotics).

There are many related topics that deal with modeling the real world. Obtaining a shared understanding of a domain under investigation is an essential part of Information Systems design, for which conceptual modeling plays a crucial role. A sound ontological background is needed to assess the clarity and understandability of the domain modeling task. The more complex the domain, the more a sound ontological commitment is needed to determine the precise set of concepts that are relevant for that domain. For example, understanding the Human Genome is probably one of the biggest challenges faced by humans. Deciphering the language of life requires complex conceptual modeling work to conceptually characterize the basic building units of such a language of life. The conceptual model of the human genome needs a precise ontological commitment that identifies its main conceptual notions together with their various set of relationships. An example can be found in the work of Pastor and colleagues [79], [24], [24], [61], and Ceri and colleagues [6], [4]. For the past decade, they have generated a continuous evolution of the modeling of the human genome, followed by a characterization of its ontological commitment. This shows an interesting attempt to prove that ontology-driven conceptual modeling improves conventional conceptual model creation and understanding.

Future research topics will continue to focus on using ontologies and ontological analysis to model the real world. Rich representations of the domain are needed, particularly from the point of view of systems. Lukyanenko, Storey, and Pastor [49], for example, propose a systemist ontology, the Bunge's Systemic Ontology (BSO). This is also an attempt to "make a machine understand" in AI, which requires a rigorous conceptualization. Of course, this requires much clarity in the representation of the involved concepts. Other future research topics include differentiating the roles of ontology [31]; conceptual modeling validation and learning; primitives [66], [67]; complexity management [25]; and semantic interoperability [32].

3. Overview of the Special Issue Papers from ER'18

The 37th International Conference on Conceptual Modeling served as a forum where novel areas such as Big Data and Artificial Intelligence as well as their fundamental and theoretical issues that are directly related to conceptual modeling were discussed. In this ER'18 edition, we placed special emphasis on research focused on machine and deep learning and how conceptual modeling can be successfully be applied soon in these areas in order to increase the success rates of the application of these Artificial Intelligence techniques [78].

For the 37th conference, 151 full papers were submitted. Each paper was reviewed by at least three reviewers and, based upon these reviews, 30 full papers and 13 short papers were selected for publication in the proceedings and presentation at the conference. The acceptance rate for regular papers was 19.87%, and for regular and short papers together, 28.48%. These papers were organized into 13 sessions that represent leading research areas in conceptual modeling, including topics related to fundamentals of conceptual modeling, ontologies, semi-structured and spatio-temporal modeling, language and models, and conceptual modeling for machine learning. Moreover, 7 high quality workshops were conducted in the conference for more specific and concrete research topics based on conceptual modeling.

The six papers included in this special issue were selected based on the scores these papers received during the reviewing process, as well as the quality and maturity of the research. Each paper is an extended and carefully revised version of the original paper from the conference. All the papers went through a rigorous reviewing process before they were accepted for this special issue.

The first paper, *Multi-Level Conceptual Modeling: Theory, Language and Application* by Fonseca et al. is a contribution in the area of multi-level conceptual modeling. In a concrete way, the authors focus on the problem of the representation needs of metaclasses that classify multiple instances that are themselves types in different subject domains. The authors address this issue by proposing an expressive multi-level conceptual modeling language (dubbed ML2). They show that ML2 enables the expression of a number of multi-level modeling scenarios that cannot be currently expressed in existing multi-level modeling languages. A textual syntax for ML2 is provided with an implementation in Xtext. They also discuss how the formal theory influences the language in two aspects: (i) by providing rigorous justification for the language's syntactic rules, which follow MLT (Multi Level Theory) theorems; and (ii) by forming the basis for model simulation and verification. Finally, the authors demonstrate that the language's practical relevance can reveal problems in multi-level taxonomic structures by using Wikidata fragments.

The second paper, *Seamless Conceptual Modeling of Processes with Transactional and Analytical Data* by Combi et al. investigates the connection between processes and the data they generate for further analysis from a conceptual modeling perspective. The authors focus on data- and decision-intensive contexts and start by arguing that business process activities need to access the data stored both in databases and data warehouses. They propose a novel conceptual view that bridges process activities and data, which allows the designer to model the connection between business processes and database models and define the operations to perform. This provides interesting insights into the overall connected perspective and suggestions for identifying activities that are crucial for decision support.

The third paper, *Managing Polyglot Systems Metadata with Hypergraphs* by Hewasinghage et al. addresses the task of the heterogeneity of the data storage models in NoSQL world. The authors propose a hypergraph-based approach for representing the catalog of data storage metadata in a polyglot system. Starting from an existing common programming interface to NoSQL systems, they extend and formalize it as hypergraphs. This formalization is based on the definition of design constraints and query transformation rules for three representative data store types. Moreover, they propose a simple query rewriting algorithm using the catalog itself for these data store types and provide a prototype implementation.

The fourth paper, *Types and Taxonomic Structures in Conceptual Modeling: A Novel Ontological Theory and Engineering Support* by Guizzardi et al. contributes to the area of conceptual modeling and ontologies. The authors show that, even though types are fundamental for conceptual modeling and knowledge representation, there is still a lack of theoretical foundation for properly addressing the definition of types. Thus, the authors start by revising the theory of types in the UFO (Unified Foundational Ontology). In this extended paper, the authors propose OntoUML 2, as a new version of their formerly proposed Conceptual Model, *OntoUML*. The new proposed formal theory is employed to support the definition of UFO-based lightweight semantic web ontologies with ontological constraint checking in OWL. Another significant contribution is the empirical evidence provided from different areas, such as psychology or linguistics.

The fifth paper, *An Ontological Analysis of Software System Anomalies and Their Associated Risks* by Duarte et al. investigates the anomalies of software systems from an ontological perspective. Their work is based on the notion that, in the design of software systems, there are many different conceptual modeling approaches having different vocabulary, terms, and notations, thereby making the maintenance and solving of software anomalies difficult. Therefore, in an attempt to deal with this heterogeneity, the authors propose two reference conceptual models: (i) an Ontology of Software Defects, Errors and Failures (OSDEF), which takes into account an ecosystem of software artifacts; and (ii) a Reference Ontology of Software Systems (ROSS), which characterizes software systems and related artifacts at different levels of abstraction. Finally, these two conceptual models are put into practice under a value and risk point-of-view, by integrating them with the Common Ontology of Value and Risk (COVR).

The sixth paper, *Analysis and Evaluation of Document-oriented Structures* by Gómez et al. addresses the challenge of data structuring alternatives in document-oriented systems. The authors first propose a semi-automatic generation of many suitable alternatives for data structuring, given an initial UML model. Then, the authors propose a set of metrics that allows designers to compare different alternatives for JSON compatible schemas, and therefore, assist in the decision criteria for schema analysis and design process. Finally, the authors propose a validation scenario as a matter of guide on how to use the proposed model and its metrics in a schema recommendation perspective.

4. Beyond ER'18

There were many other topics presented in the conference that could not be included in a limited special issue. However, they provide motivation and foundations for continued research on conceptual modeling.

We hope you enjoy this special issue and appreciate the work of the contributing authors.

Acknowledgements

We wish to thank the former Editor-in-Chief of the *Data and Knowledge Engineering (DKE)* journal, Dr. Peter Chen, for his strong support for this special issue and all the conceptual modeling conferences year after year. We would also like to thank the new Editor-in Chief of the *DKE*, Dr. Carson Woo, for his support of this special issue. We are grateful to our excellent group of reviewers who carried out the review process in a timely manner while still meeting the high expectations of a scholarly journal.

We are thankful to all the authors of these extended papers who made an excellent effort in extending their conference papers from the ER'18 conference. Without them, this special issue would not have been possible at all.

Thank you to Anna Bernasconi, Giancarlo Guizzardi, Stephen Liddle, Roman Lukyanenko and Oscar Pastor for their comments on a prior version of this paper.

The research reported in this paper was partially funded by the ECLIPSE-UA (RTI2018-094283-B-C32) project funded by Spanish Ministry of Science, Innovation, and Universities.

Guest editors

Juan Trujillo
University of Alicante, Spain
Karen C. Davis
Miami University, USA
Xiaoyong Du
Renmin University of China, China
Veda C. Storey
Georgia State University, Atlanta, USA
Ernesto Damiani
Università degli Studi di Milano, Italy

References

[1] Claudio Agostino Ardagna et al. "A Model-Driven Methodology for Big Data Analytics-as-a-Service". In: *2017 IEEE International Congress on Big Data (BigData Congress)*. Ed. by George Karypis and Jia Zhang. IEEE, June 2017, pp. 105–112.

[2] Sylvio Barbon Junior et al. "Evaluating Trace Encoding Methods in Process Mining". In: *From Data to Models and Back*. Ed. by Juliana Bowles, Giovanna Broccia, and Mirco Nanni. Cham: Springer International Publishing, 2021, pp. 174–189.

[3] Nabila Berkani et al. "The contribution of linked open data to augment a traditional data warehouse". In: *Journal of Intelligent Information Systems* (2020), pp. 1–25.

[4] Anna Bernasconi, Arif Canakoglu, and Stefano Ceri. "From a conceptual model to a knowledge graph for genomic datasets". In: *International Conference on Conceptual Modeling*. Cham: Springer International Publishing, 2019, pp. 352–360.

[5] Anna Bernasconi, Arif Canakoglu, Stefano Ceri, et al. "Ontology-driven metadata enrichment for genomic datasets". In: CEUR Workshop Proceedings 2275 (2018). Ed. by Christopher J. O. Baker et al.

[6] Anna Bernasconi et al. "Conceptual modeling for genomics: building an integrated repository of open data". In: *International Conference on Conceptual Modeling*. Springer. 2017, pp. 325–339.

[7] Mario Bunge. *Treatise on basic philosophy: Ontology I: the furniture of the world*. Vol. 3. Springer Science & Business Media, 1977.

[8] Andrew Burton-Jones et al. "A semiotic metrics suite for assessing the quality of ontologies". In: *Data & Knowledge Engineering* 55.1 (2005), pp. 84–102.

[9] Diego Calvanese et al. "Optique: OBDA solution for big data". In: *Extended Semantic Web Conference*. Springer. 2013, pp. 293–295.

[10] Hsinchun Chen, Roger HL Chiang, and Veda C Storey. "Business intelligence and analytics: From big data to big impact". In: *MIS quarterly* (2012), pp. 1165–1188.

[11] Peter Pin-Shan Chen. "The entity-relationship model—toward a unified view of data". In: *ACM Transactions on Database Systems (TODS)* 1.1 (Mar. 1976), pp. 9–36.

[12] Jordi Conesa, Veda C Storey, and Vijayan Sugumaran. "Usability of upper level ontologies: The case of ResearchCyc". In: *Data & Knowledge Engineering* 69.4 (2010), pp. 343–356.

[13] Houda Daki et al. "Big Data management in smart grid: concepts, requirements and implementation". In: *Journal of Big Data* 4.1 (Dec. 2017), p. 13.

[14] Ernesto Damiani et al. "Toward Model-Based Big Data-as-a-Service: The TOREADOR Approach". In: *Advances in Databases and Information Systems - 21st European Conference (ADBIS)*. Ed. by Marite Kirikova, Kjetil Nørvåg, and George A Papadopoulos. Vol. 10509. Lecture Notes in Computer Science. Springer, 2017, pp. 3–9.

[15] Lois M. L. Delcambre et al. "A Reference Framework for Conceptual Modeling". In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 27–42.

[16] Francesco Di Tria, Ezio Lefons, and Filippo Tangorra. "Evaluation of Data Warehouse Design Methodologies in the Context of Big Data". In: *Big Data Analytics and Knowledge Discovery - 19th International Conference, DaWaK 2017, Lyon, France, August 28-31, 2017, Proceedings*. Ed. by Ladjel Bellatreche and Sharma Chakravarthy. Vol. 10440. Lecture Notes in Computer Science. Springer, 2017, pp. 3–18.

[17] Francesco Di Tria, Ezio Lefons, and Filippo Tangorra. "Evaluation of Data Warehouse Design Methodologies in the Context of Big Data". In: *Journal of Big Data*. Ed. by Daniela E Damian, Anna Perini, and Seok-Won Lee. 3rd ed. Vol. 11157. Lecture Notes in Computer Science 2. Indianapolis, IN: Springer, 2017, pp. 3–18.

[18] Humaira Ehsan, Mohamed A Sharaf, and Panos K Chrysanthis. "Efficient Recommendation of Aggregate Data Visualizations". In: *IEEE Transactions on Knowledge and Data Engineering* 30.2 (Feb. 2018), pp. 263–277.

[19] David W Embley and Stephen W Liddle. "Big data—conceptual modeling to the rescue". In: *International Conference on Conceptual Modeling*. Springer. 2013, pp. 1–8.

[20] David W Embley et al. "Ontology-based extraction and structuring of information from data-rich unstructured documents". In: *Proceedings of the seventh international conference on Information and knowledge management*. 1998, pp. 52–59.

[21] Hanif Eridaputra, Bayu Hendradjaya, and Wikan Danar Sunindyo. "Modeling the requirements for big data application using goal oriented approach". In: *2014 International Conference on Data and Software Engineering (ICODSE)*. IEEE, Nov. 2014, pp. 1–6.

[22] Owen Eriksson and Pär J Ågerfalk. "Speaking things into existence: ontological foundations of identity representation and management". In: *Information Systems Journal* (2021).

[23] Christina Feilmayr and Wolfram Wöß. "An analysis of ontologies and their success factors for application to business". In: *Data & Knowledge Engineering* 101 (2016), pp. 1–23.

[24] Ana Ma Martınez Ferrandis, Oscar Pastor López, and Giancarlo Guizzardi. "Applying the principles of an ontology-based approach to a conceptual schema of human genome". In: *International Conference on Conceptual Modeling*. Springer. 2013, pp. 471–478.

[25] Guylerme Figueiredo et al. "Breaking into pieces: An ontological approach to conceptual model complexity management". In: *2018 12th International Conference on Research Challenges in Information Science (RCIS)*. IEEE. 2018, pp. 1–10.

[26] Vittorio Gallese and George Lakoff. "The brain's concepts: The role of the sensory-motor system in conceptual knowledge". In: *Cognitive neuropsychology* 22.3-4 (2005), pp. 455–479.

[27] Matteo Golfarelli and Stefano Rizzi. "A model-driven approach to automate data visualization in big data analytics". In: *Information Visualization* 19.1 (Jan. 2020), pp. 24–47.

[28] Matteo Golfarellli, Daario Maio, and Stefano Rizzi. "The Dimensional Fact Model: A Conceptual Model for Data Warehouses". In: *International Journal of Cooperative Information Systems* 07.02n03 (June 1998), pp. 215–247.

[29] Nicola Guarino. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. Vol. 46. IOS press, 1998.

[30] Nicola Guarino, Giancarlo Guizzardi, and John Mylopoulos. "On the philosophical foundations of conceptual models". In: *Information Modelling and Knowledge Bases* 31.321 (2020), p. 1.

[31] Giancarlo Guizzardi. “On ontology, ontologies, conceptualizations, modeling languages, and (meta) models”. In: *Frontiers in artificial intelligence and applications* 155 (2007), p. 18.

[32] Giancarlo Guizzardi. “Ontology, ontologies and the “I” of FAIR”. In: *Data Intelligence* 2.1-2 (2020), pp. 181–191.

[33] Giancarlo Guizzardi and Gerd Wagner. “Some applications of a unified foundational ontology in business modeling”. In: *Business systems analysis with ontologies*. IGI Global, 2005, pp. 345–367.

[34] Giancarlo Guizzardi et al. “Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story”. In: *Applied ontology* 10.3-4 (2015), pp. 259–271.

[35] Long Guo et al. “CRAN: A Hybrid CNN-RNN Attention-Based Model for Text Classification”. In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 571–585.

[36] Fuyuki Ishikawa. “Concepts in Quality Assessment for Machine Learning - From Test Data to Arguments”. In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 536–544.

[37] Zongze Jin et al. “Leveraging the Dynamic Changes from Items to Improve Recommendation”. In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 507–520.

[38] Paul Johannesson and Erik Perjons. “An Ontological Analysis of the Notion of Treatment”. In: *International Conference on Conceptual Modeling*. Springer. 2020, pp. 303–314.

[39] S. B. Junior et al. “Anomaly Detection on Event Logs with a Scarcity of Labels”. In: *2020 2nd International Conference on Process Mining (ICPM)*. 2020, pp. 161–168.

[40] Ralph Kimball and Margy Ross. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd ed. Indianapolis, IN: Wiley, 2013. ISBN: 978-1-118-53080-1.

[41] Natalija Kozmina, Laila Niedrite, and Janis Zemnickis. “Information Requirements for Big Data Projects: A Review of State-of-the-Art Approaches”. In: *Databases and Information Systems - 13th International Baltic Conference, DB&IS 2018, Trakai, Lithuania, July 1-4, 2018, Proceedings*. Ed. by Audrone Lupeikiene, Olegas Vasilecas, and Gintautas Dzemyda. Vol. 838. Communications in Computer and Information Science. Springer, 2018, pp. 73–89.

[42] Ana Lavalle, Alejandro Maté, and Juan Trujillo. "An Approach to Automatically Detect and Visualize Bias in Data Analytics". In: *Proceedings of the 22nd International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP)*. Ed. by Il-Yeol Song, Katja Hose, and Oscar Romero. Vol. 2572. {CEUR} Workshop Proceedings. CEUR-WS.org, 2020, pp. 84–88.

[43] Ana Lavalle, Alejandro Maté, and Juan Trujillo. "Requirements-Driven Visualizations for Big Data Analytics: A Model-Driven Approach". In: *International Conference on Conceptual Modeling (ER)*. Lecture No. Springer-Verlag, 2019, pp. 78–92.

[44] Ana Lavalle et al. "Visualization Requirements for Business Intelligence Analytics: A Goal-Based, Iterative Framework". In: *2019 IEEE 27th International Requirements Engineering Conference (RE)*. Ed. by Daniela E Damian, Anna Perini, and Seok-Won Lee. IEEE, Sept. 2019, pp. 109–119.

[45] Yeting Li et al. "Learning Restricted Regular Expressions with Interleaving from XML Data". In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 586–593.

[46] Lucas Henrique C Lima et al. "An analysis of the collaboration network of the International Conference on Conceptual Modeling at the Age of 40". In: *Data & Knowledge Engineering* 130 (2020), p. 101866.

[47] Lingting Lin, Chen Lin, and Yongxuan Lai. "Realtime Event Summarization from Tweets with Inconsistency Detection". In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 555–570.

[48] Sergio Luján-Mora, Juan Trujillo, and Il-Yeol Song. "A UML profile for multidimensional modeling in data warehouses". In: *Data & Knowledge Engineering (DKE)* 59.3 (Dec. 2006), pp. 725–769.

[49] Roman Lukyanenko, Veda C Storey, and Oscar Pastor. "Foundations of information technology based on Bunge's systemist philosophy of reality". In: *Software and Systems Modeling* (2021), pp. 1–18.

[50] Roman Lukyanenko et al. "Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content". In: *MIS Quarterly* 43.2 (2019), pp. 623–648.

[51] Wolfgang Maass and Iaroslav Shcherbatyi. "Inductive Discovery by Machine Learning for Identification of Structural Models". In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 545–552.

[52] Wolfgang Maass and Veda C Storey. "Pairing Conceptual Modeling with Machine Learning". In: *Working Paper* (2021).

[53] Wolfgang Maass et al. "Data-driven meets theory-driven research in the era of big data: opportunities and challenges for information systems research". In: *Journal of the Association for Information Systems* 19.12 (2018), p. 1.

[54] Jose-Norberto Mazón and Juan Trujillo. "An MDA approach for the development of data warehouses". In: *Decision Support Systems (DSS)* 45.1 (Apr. 2008), pp. 41–58.

[55] Melinda McDaniel and Veda C Storey. "Evaluating domain ontologies: clarification, classification, and challenges". In: *ACM Computing Surveys (CSUR)* 52.4 (2019), pp. 1–44.

[56] Melinda McDaniel, Veda C Storey, and Vijayan Sugumaran. "Assessing the quality of domain ontologies: Metrics and an automated ranking system". In: *Data & Knowledge Engineering* 115 (2018), pp. 32–47.

[57] Sergio Orenga-Roglá and Ricardo Chalmeta. "Methodology for the Implementation of Knowledge Management Systems 2.0". In: *Business & Information Systems Engineering (BISE)* 61.2 (Apr. 2019), pp. 195–213.

[58] Grace Park et al. "IRIS: A goal-oriented big data analytics framework on Spark for better Business decisions". In: *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, Feb. 2017, pp. 76–83.

[59] Jeffrey Parsons and Yair Wand. "Choosing classes in conceptual modeling". In: *Communications of the ACM* 40.6 (1997), pp. 63–69.

[60] Jeffrey Parsons and Yair Wand. "Emancipating instances from the tyranny of classes in information modeling". In: *ACM Transactions on Database Systems (TODS)* 25.2 (2000), pp. 228–268.

[61] Óscar Pastor et al. "Using conceptual modeling to improve genome data management". In: *Briefings in Bioinformatics* (2020).

[62] Oscar Peña, Unai Aguilera, and Diego López-de-Ipiña. "Exploring LOD through metadata extraction and data-driven visualizations". In: *Program* 50.3 (July 2016), pp. 270–287.

[63] Sandeep Purao and Veda C Storey. "A multi-layered ontology for comparing relationship semantics in conceptual models of databases". In: *Applied Ontology* 1.1 (2005), pp. 117–139.

[64] Jan Recker et al. "From representation to mediation: a new agenda for conceptual modeling research in a digital world". In: *MIS Quarterly* (2021).

[65] Michael Rosemann, Peter Green, and Marta Indulska. "A reference methodology for conducting ontological analyses". In: *International Conference on Conceptual Modeling*. Springer. 2004, pp. 110–121.

[66] Fabiano B Ruy et al. "From reference ontologies to ontology patterns and back". In: *Data & Knowledge Engineering* 109 (2017), pp. 41–69.

[67] Tiago Prince Sales and Giancarlo Guizzardi. "Ontological anti-patterns: Empirically uncovered error-prone structures in ontology-driven conceptual models". In: *Data & Knowledge Engineering* 99 (2015), pp. 72–104.

[68] Tiago Prince Sales et al. "An ontological analysis of value propositions". In: *2017 IEEE 21st International Enterprise Distributed Object Computing Conference (EDOC)*. IEEE. 2017, pp. 184–193.

[69] Veda C Storey. "Conceptual modeling meets domain ontology development: a reconciliation". In: *Journal of Database Management (JDM)* 28.1 (2017), pp. 18–30.

[70] Veda C Storey, Robert C Goldstein, and Harald Ullrich. "Naive semantics to support automated database design". In: *IEEE Transactions on Knowledge and data engineering* 14.1 (2002), pp. 1–12.

[71] Veda C Storey and Sandeep Purao. "Understanding relationships: Classifying verb phrase semantics". In: *International Conference on Conceptual Modeling*. Springer. 2004, pp. 336–347.

[72] Veda C. Storey, Juan C. Trujillo, and Stephen W. Liddle. "Research on conceptual modeling: Themes, topics, and introduction to the special issue". In: *Data & Knowledge Engineering (DKE)* 98 (July 2015), pp. 1–7.

[73] Vijayan Sugumaran and Veda C Storey. "Ontologies for conceptual modeling: their creation, use, and management". In: *Data & knowledge engineering* 42.3 (2002), pp. 251–271.

[74] Vijayan Sugumaran and Veda C Storey. "The role of domain ontologies in database design: An ontology management and conceptual modeling environment". In: *ACM Transactions on Database Systems (TODS)* 31.3 (2006), pp. 1064–1094.

[75] Roberto Tardio, Alejandro Mate, and Juan Trujillo. "An iterative methodology for big data management, analysis and visualization". In: *IEEE International Conference on Big Data (Big Data)*. IEEE, Oct. 2015, pp. 545–550.

[76] Roberto Tardio, Alejandro Mate, and Juan Trujillo. "An Iterative Methodology for Defining Big Data Analytics Architectures". In: *IEEE Access* 8 (2020), pp. 210597–210616.

[77] J. Trujillo et al. "Designing data warehouses with OO conceptual models". In: *Computer* 34.12 (2001), pp. 66–75.

[78] Juan C. Trujillo et al., eds. *Conceptual Modeling - 37th International Conference, ER 2018*. Vol. 11157. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018.

[79] Michaël Verdonck et al. "Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study". In: *Information Systems* 81 (2019), pp. 92–103.

[80] Yair Wand and Ron Weber. “An ontological model of an information system”. In: *IEEE transactions on software engineering* 16.11 (1990), pp. 1282–1292.

[81] Yair Wand and Ron Weber. “On the deep structure of information systems”. In: *Information Systems Journal* 5.3 (1995), pp. 203–223.

[82] Yair Wand et al. “Theoretical foundations for conceptual modelling in information systems development”. In: *Decision support systems* 15.4 (1995), pp. 285–304.

[83] Xuan Wang et al. “Mining Rules with Constants from Large Scale Knowledge Bases”. In: *Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings*. Ed. by Juan Trujillo et al. Vol. 11157. Lecture Notes in Computer Science. Springer, 2018, pp. 521–535.

[84] Ron Weber. “Conceptual modelling and ontology: Possibilities and pitfalls”. In: *Journal of Database Management (JDM)* 14.3 (2003), pp. 1–20.

[85] Carson Woo. “The role of conceptual modeling in managing and changing the business”. In: *International Conference on Conceptual Modeling*. Springer. 2011, pp. 1–12.

[86] Xuesong Zhao. “A Study on the Applications of Big Data in Cross-Border E-Commerce”. In: *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*. IEEE, Oct. 2018, pp. 280–284.