



**HAL**  
open science

## A probabilistic digital twin model for inland waterway transportation systems using Bayesian networks

E Alexandra Micu, Fereshteh Asgari, Mostepha Khouadjia, Lorcan Connolly,  
Kostas Zavitsas

### ► To cite this version:

E Alexandra Micu, Fereshteh Asgari, Mostepha Khouadjia, Lorcan Connolly, Kostas Zavitsas. A probabilistic digital twin model for inland waterway transportation systems using Bayesian networks. 7th Conference on Sustainable Mobility (CSuM), Sep 2024, Karditsa, Greece. hal-04686200

**HAL Id: hal-04686200**

**<https://hal.science/hal-04686200>**

Submitted on 3 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A probabilistic digital twin model for inland waterway transportation systems using Bayesian networks

E. Alexandra Micu<sup>1</sup>, Fereshteh Asgari<sup>2</sup>, Mostepha Khouadjia<sup>2</sup>, Lorcan Connolly<sup>1</sup>,  
Kostas Zavitsas<sup>3</sup>

<sup>1</sup> Research Driven Solutions, 1a Saint Kevin's Avenue, Blackpitts, Dublin, D08TX29, Ireland

<sup>2</sup> Technological Research Institute- IRTY SystemX, Paris-Saclay, 2 Bd Thomas Gobert, 91120  
Palaiseau, France

<sup>3</sup> VLTN, De Keyserlei 58-60 bus 19, 2018 Antwerp | Belgium

`alexandra.micu@researchdrivensolutions.ie`

**Abstract.** Resilience-centric Smart, Green, Networked EU Inland Waterways (ReNEW) is a Horizons Europe project that aims to address IWT systems' dynamicity, heterogeneity, and complexity as well as to develop strategies for making IWT systems smarter, greener, more sustainable, and climate-resilient. A Digital Twin (DT) is a powerful tool for modelling the complexity and interdependencies of complex systems, such as IWT systems. Data in digital twins can be represented by knowledge graphs (KG), which efficiently store information in a structured way and capture the complexity of the real world. Ontology-based knowledge graphs capture all entities within IWT systems hierarchically, allowing humans and machines to easily interpret and analyse the data. This paper introduces the concept of creating a probabilistic digital twin using Bayesian networks to account for uncertainty and forecast how the IWT system will respond to predicted and unexpected events. A Bayesian Network is a probabilistic graphical representation that combines expert belief alongside sensor data to model the relationship between the system's entities. As the evidence changes, the model is updated to provide more accurate results. Risk and safety analyses are then performed to assess the reliability of IWT infrastructure.

**Keywords:** Bayesian Networks, Digital Twins, Inland Waterways Transport, Knowledge Graph, Ontology, Resilience, Sustainability, ReNEW.

## 1 Introduction

Climate change has already caused a substantial impact on IWT infrastructure [1], and it is anticipated that this trend will continue and worsen [2]. People, businesses, and local and regional economies may be negatively affected by disruptions of IWT [3-7]. Climate-neutral and climate-resilient solutions must be planned, assessed, and implemented for IWT to function effectively [8].

IWT systems can be made more resilient by taking into account the complexity of their components and the interdependencies among them. It is becoming increasingly evident that digitalization is enabling sustainable transportation solutions, which have the potential to enhance resilience and provide economic, environmental, and societal benefits. The use of digital twins (DTs) for the modelling of complex systems, such as IWT systems, can be extremely useful, especially when predicting their reactions to events that are either predicted or unpredicted [10]. Through DTs, processes and operations can be optimized in real-time, resulting in better results than standard methodologies [11]. A digital twin model can be created using either physics or statistics.

A Bayesian Network (BN) is employed in this paper to generate a probabilistic twin to present the interdependencies between the IWT infrastructure components in a mathematical manner. The Bayesian network approach is used in DT modelling to address uncertainty introduced by unknown or missing data as well as uncertainty imposed on the model itself. The BN model has the advantage of being able to be updated as new (field) records are made available, resulting in more accurate results [12-13].

## **2 Resilience-centric Smart, Green, Networked EU Inland Waterways**

### **2.1 Project introduction**

Resilience-centric Smart, Green, Networked EU Inland Waterways (ReNEW) project, which includes 24 participants from 11 countries in the European Union, aims to assist inland waterway transportation system in transitioning to a more sustainable and climate-resilient sector. As part of the project, previous results will be leveraged, collaborations will be formed, and the following outputs will be developed:

- Decision support framework for IWT resilience and sustainability that integrates innovative models of network interdependencies, probabilistic risk analysis, and resilience quantification, identifying short- and long-term initiatives enhancing resilience.
- Innovations in infrastructure resilience and sustainability built on autonomy and green energy solutions.
- IWT data space and generic digital twin for primarily sharing infrastructure monitoring, transport management, and emergency systems and climate solutions.
- Living Labs addressing specifically inland waterway resilience and focusing on integrated IW and hinterland infrastructure.

### **2.2 Methodology for EU resilient and sustainable IWT infrastructure system**

ReNEW aims at developing the methodological framework for strategy planning for resilient and sustainable green IWT infrastructure. In addition, it will develop a toolbox of tools and solutions to analyse the interdependencies and safety risks of infrastructure,

the reconfiguration of transport services, and the impact of natural and human-caused disasters on the ecological environment. Lastly, it will identify a baseline building block, model and tool required to guide, monitor, and evaluate Living Labs' (LLs) interventions.

### **3 Methodology**

IWT systems can be modelled with the help of a Digital Twin (DT) in order to gain a more comprehensive understanding of the complexity of those systems and their interdependencies. Knowledge graphs can be used to represent data in digital twins as they effectively store information in a structured manner and capture the complexity of the real world. IWT systems can be interpreted and analysed more easily using ontology-based knowledge graphs, which organize all entities within the system hierarchically.

#### **3.1 Digital Twin**

Digital twins describes a physical system using a digital model that includes both recorded and up-to-date data about the system. The information exchange between the physical system and its digital model takes place is automated. A DT has a variety of applications such as classification, prediction, optimisation etc. A digital twin can be constructed for one or more elements (such as lock chambers), subsystems (such as locks), systems (such as an IWT infrastructure), or networks (such as an entire IWT network).

Modelling digital twins can be based on a number of approaches, including physics, statistics, or a combination of the two. Statistical techniques such as Bayesian networks can be used to graphically model the variables of complex systems and their conditional dependencies. The use of Bayesian networks in digital twin modelling offers a good way of addressing both the uncertainty introduced by unknown or missing data as well as the uncertainty associated with the model itself.

#### **3.2 Ontology**

An ontology is a formal explicit specification of a shared conceptualization of a domain of interest [14]. With an ontology we can model a system through a set of concepts/things and the properties that are used to describe them. The Ontology is completed by defining the underlying relationships between the concepts. An ontology together with a set of individual instances of classes constitutes a knowledge base and offers services to facilitate interoperability across multiple and heterogeneous systems [15].

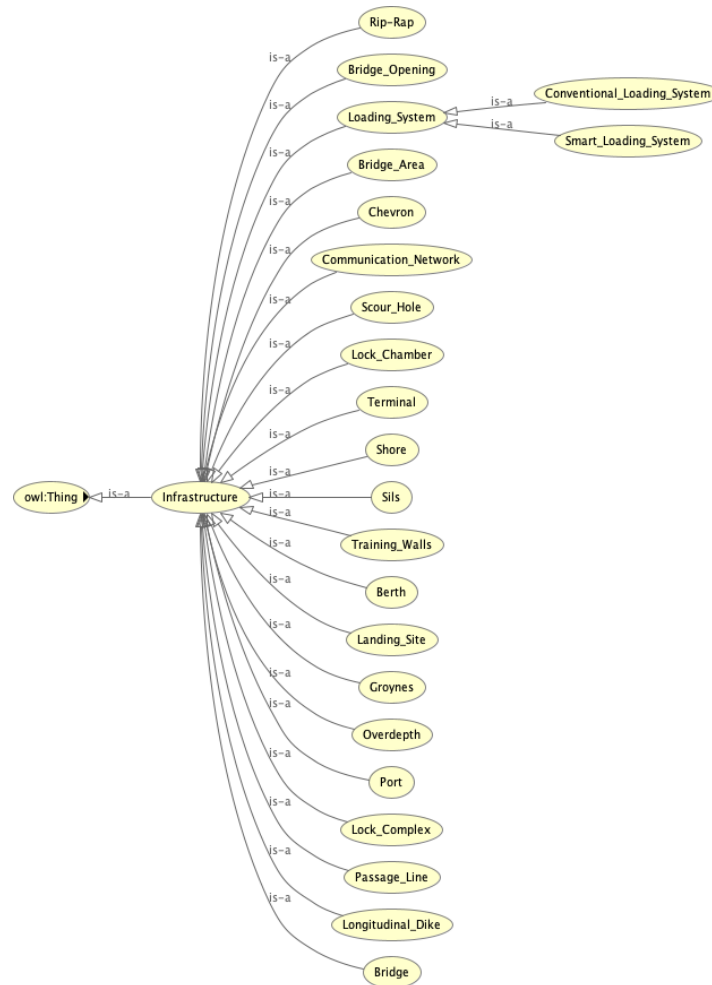
In modelling the IWT ontology, to the author's knowledge there was no existing ontology for IWT systems. Hence, the ontology was conceptualized from scratch. In this process we also considered the elements of IWT network resource, and the data model presented in the EuRIS portal [16] (as of a bottom-up approach). The EuRIS portal provides all waterway and traffic related information of 13 European countries.

This data contains the data related to IWT network infrastructures such as waterways, terminals, bridges, and locks.

Each of these entities, composed of a set of properties, for example a fairway section as an entity, has properties such as waterway-name (to which waterway it belongs), fairway reference code, fairway sequence number, coordinate as a line string (a set of coordinates that present the whole fairway section, its length, width and depth).

Based on these elements, our IWT network ontology contains 124 classes, 42 relations and 143 attributes. We used the Protégé application [17] for the conceptualization of this ontology. An overview of the ontology for the infrastructure is shown in Fig. 1.

An ontology provides the foundation for formal semantics in a knowledge graph. It is possible to think of them as the data schema of the KG. In essence, they serve as the common language for KG creators and users to communicate the meaning of the data contained within.



**Fig. 1.** IWT Ontology: Infrastructure overview.

### 3.3 Knowledge Graphs

Knowledge graphs (KGs) are graphical representations of knowledge in the real world [18]. In a KG, the data is depicted as nodes or vertices. The relationships between data are described by edges. Data entities and their relationships represent the key components of a KG [19]. Data entities comprises (in relation to IWT networks) personnel, events, sensors, regulations, vessels, infrastructure, waterway, etc. Example of relationships in a KG include “process”, “part of”, “located in”, “is component of”, “type”, “similar to”, “country”, “width of”, “length of”, “depth of”, “height of”, “included in”, “depends on”, etc. A ReNEW IWT sample knowledge graph is shown in Fig. 2.

The core of the knowledge graph is the knowledge model, which consists of many interconnected descriptions of concepts, entities, relationships, and events. Having a formal semantic framework allows people and computers to process descriptions efficiently and clearly. Therefore, knowledge graphs can serve as the foundation for building Bayesian network models for probabilistic reasoning.

In the context of the ReNEW project, the KG is extended to include IWT resilience indicators. The indicators proposed for the IWT capture both the performance metrics associated to single network infrastructure components, as well as the IWT system's serviceability performance. To illustrate the differentiation between the two levels, navigability is assessed in the context of individual fairways, where the fairway is not navigable above or below some predefined water level and flow thresholds. At a IWT system level, corridor navigability is established, and the throughput potential of the network is assessed.

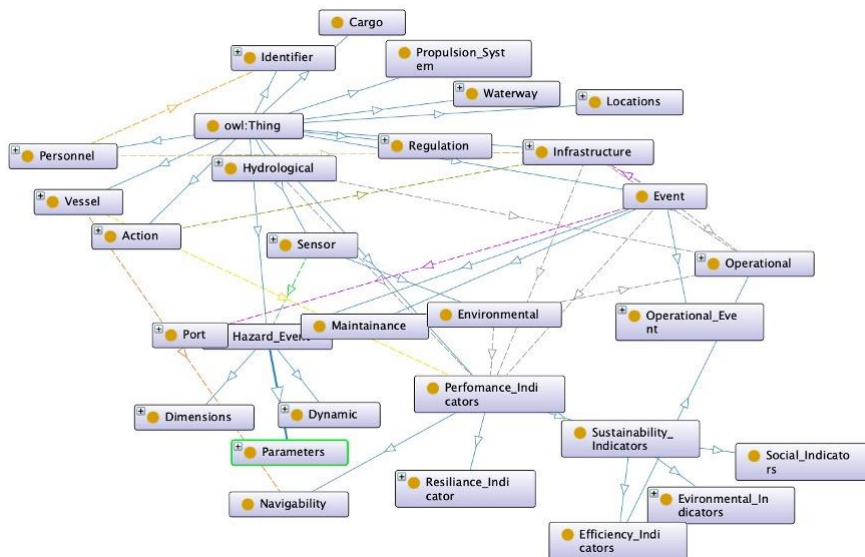
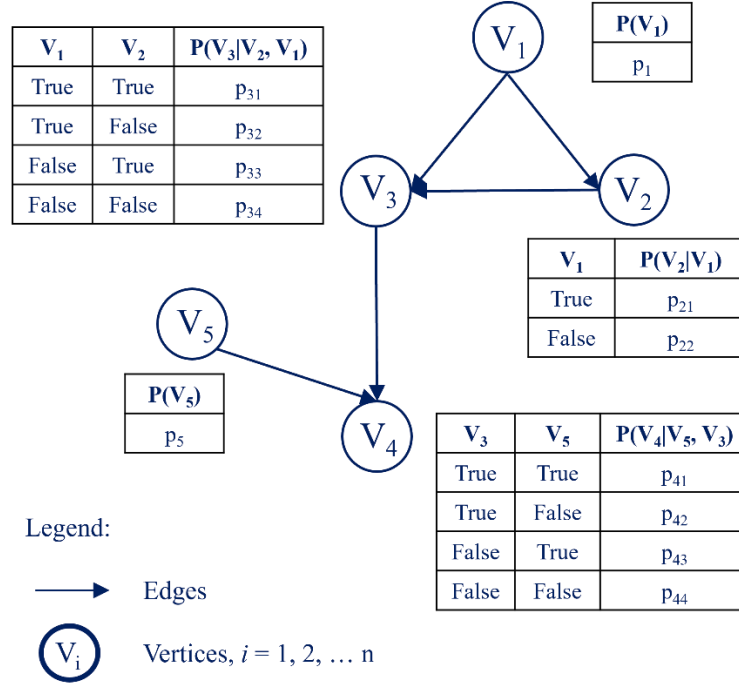


Fig. 2. ReNEW IWT sample knowledge graph.

### 3.4 Bayesian Networks

A Bayesian Network is a probabilistic tool that helps to show the connections between data, ideas, information, and concepts within a given system. An example of a BN model is illustrated in Figure 3. A Bayesian network consists of nodes/vertices for representing objects, variables, or elements in a system, and edges for representing their relationships. A given vertex in the system exists in a particular number of states, which indicate its current status. Discrete conditional probabilities for each vertex are displayed in the conditional probability tables (CPTs).



**Fig. 3.** A notional Bayesian Network model.

A Bayesian network is based on Bayes's Theorem, as shown in Equation 1. The Bayes' theorem defines prior probability as the probability of a variable being in a particular condition state at any given moment. A conditional probability describes the probability that a variable will be in a certain state given the state of its parents (e.g., vertices  $V_1$  and  $V_2$  in Fig. 3 are the parents of vertex  $V_3$ ). It is necessary to consider both the prior probability and the conditional probability, as well as the statistical principles that define their interdependency, in order to estimate the likelihood that a variable will be in a particular state.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

where:

$P(A|B)$  represents the posterior (or conditional) probability.

$P(A)$  represents the prior (or marginal) probability.

$P(B|A)$  represents the likelihood.

$P(B)$  represents the normalizing factor, calculated in Equation 2 based on the Total Probability Theorem:

$$P(B) = P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A}) \quad (2)$$



There are many applications of BNs, including risk, safety, and reliability evaluations, diagnostics, reasoning, causal inferences, decision-making under uncertainty, anomaly detection, automated insight, and prediction [20-21]. Modelling a BN involves the following steps. To begin with, it is necessary to identify/recognize/detect the hazards or events that can constitute a threat, as well as their characteristics (i.e., occurrence, intensity, etc). This is followed by evaluating the potential impact of hazards on critical and non-critical components (nodes) of a system. Finally, the likelihood of the network components being in various states after the hazard is measured without altering any individual nodes.

To illustrate an example of how BNs can be used to assess risk, Fig. 4 depicts the breakdown of IWT networks into systems (infrastructure, vessels, ports, etc.), subsystems (locks, bridges, etc.), and elements (lock chambers, gates, etc.). The components' current state is evaluated using a variety of data sources and expert opinions. Each damage state is assigned a set of metrics, which will be used to create CPTs. Finally, the BN model for risk assessment can then be developed. Figure 5 illustrates a notional BN model for assessing risk.

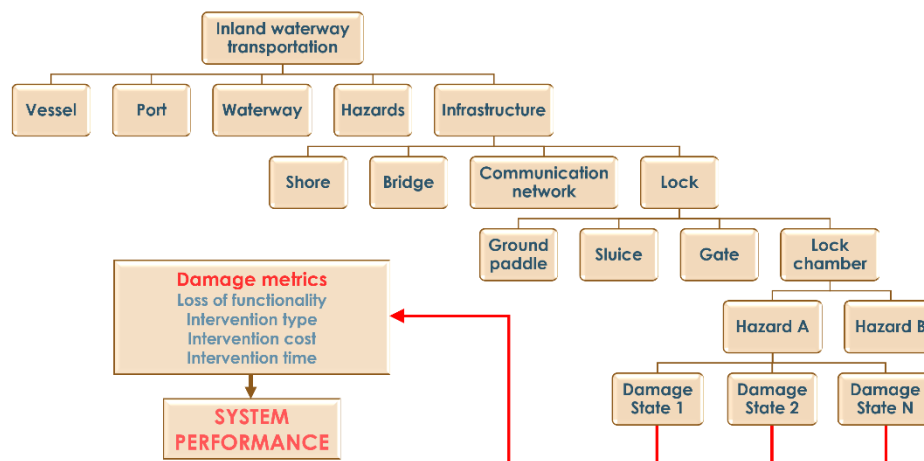


Fig. 4. IWT network breakdown.

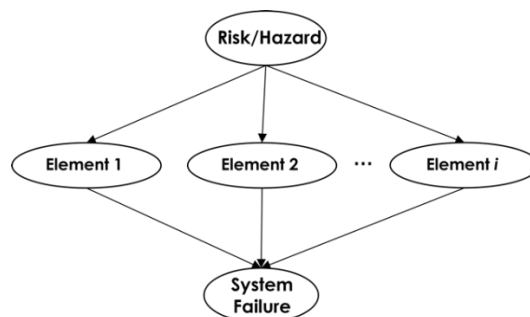


Fig. 5. A notional Bayesian Network model for risk assessment.

Bayesian networks have many applications in a wide range of fields. Bayesian networks offer numerous benefits but also some drawbacks [22]. One of the most important advantages of a BN is that it can give a visual representation of the system. They assist in making decisions when working with vast amounts of data. A BN is efficient and intuitive. BNs enable the incorporation of uncertainty into the model and the model's updating in response to new data. A BN can be used to model multiple interdependencies and help identify key components of a network. Practical restrictions, however, apply to the kind of distribution and statistical dependence that can be employed. BNs require a lot of computing power for systems that have a large number of stochastic states and variables.

### **3.5 Living Labs**

Real-world scenarios will be used to test the concepts presented in this work. Data about the IWT network will be stored in KG and then used to build BN models for each individual living lab. Bayesian networks will be used to model cause-and-effect models to determine how events related to climate change affect the City Logistics Hub's operations. To determine the navigability of the Douro River in the event of extreme weather, Bayesian networks will be used to model what-if scenarios. Lastly, with the aid of Bayesian networks, predictive models will be created that will enable the planning of mitigation strategies to encourage modal shifting.

## **4 Conclusions**

Environmental factors such as climate change have a significant impact on inland waterway transportation networks. The effectiveness of the IWT depends on smart, green, and sustainable solutions. The digitalization of sustainable transportation is associated with several benefits, including enhancing resilience, as well as benefits to local economies, the environment, and society as a whole. KGs (knowledge graphs) can be used in digital twins in order to store and represent data in a structured manner that captures the complexity of the real world. All entities within IWT systems can be represented hierarchically through ontology-based knowledge graphs, thereby facilitating intuitive interpretation and analysis by humans and machines. The paper proposes using Bayesian networks to simulate the probabilistic digital twin of complex inland waterways transportation networks. As the name implies, digital twins are digital representations of real-world objects. There is a great deal of complexity and uncertainty associated with IWT applications, which entails the use of statistical models, such as Bayesian networks. The probabilistic twin model can also be updated as new data becomes available, generating more accurate estimates over time.

## **Acknowledgements**

The ReNEW project received funding from the Horizon Europe research and innovation program under Grant Agreement No 101069682.

## References

1. Hendrickx, C., Breemersch, T.: The effect of climate change on inland waterway transport. – Transport Research Arena 2012, Papaioannou, P. (Ed.), Elsevier Science Bv, 1837–1847 (2012).
2. EEA.: Adaptation of transport to climate change in Europe – Challenges and options across transport modes and stakeholders, 60 pp (2014), DOI:10.2800/242209.
3. Jonkeren, O., Rietveld, P., van Ommeren, J.: Climate change and inland waterway transport - welfare effects of low water levels on the river Rhine. *J. Transp. Econ. Policy* 41 (3), 387–411 (2007).
4. Scholten, A. and Rothstein, B.: Navigation on the Danube - Limitations by low water levels and their impacts, JRC Technical Reports, ISBN 978-92-79-64798-7, doi:10.2788/236234, European Commission (2016).
5. Athanasatos, S., Michaelides, S., Papadakis, M.: Identification of weather trends for use as a component of risk management for port operations. – *Nat. Hazards* 72, 41–61 (2014), DOI:10.1007/s11069-012-0491-z.
6. Schwartz Jr, H., Tavasszy, L.: Transportation Resilience: Adaptation to Climate Change and Extreme Weather Events. – Transportation Research Board Conference Proceedings (2016).
7. Kwiatkowski, K.P., Chinowsky, P.S., Transportation Research, B.: Climate Change Adaptation in Transportation Infrastructure Organizations: Ranking Processes and Institutional Elements. – Transportation Research Board 96th Annual Meeting, Washington DC, United States, Transportation Research Board 17 (2017).
8. Love, G., Soares, A., Puempel, H.: Climate Change, Climate Variability and Transportation. – World Climate Conference – 3. Sivakumar, M.V.K., B.S. Nyenzi, A. Tyagi (Eds.), Elsevier Science Bv, 130–145 (2010).
9. Ammar, A., Bhatt, B., Nassereddine, H., Dadi, G.: A Blueprint for Creating Digital Twins for Transportation Assets: an Application for Highway Engineering. In Proc., 11th International Conference on Engineering Computational Technology, pp. 9-1, Civil-Comp Press, Edinburgh (2022).
10. Broo, D.G., Schooling, J.: Digital twins in infrastructure: definitions, current practices, challenges and strategies, *International Journal of Construction Management*, 1–10 (2021).
11. Ambra, T., Caris, A. and Macharis, C.: The Digital Twin concept and its role in reducing uncertainty in synchromodal transport. *Proceedings of the IPIC* (2019).
12. Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*. 1992;9:309e347.
13. Riggelsen C, Feelders A. Learning Bayesian network models from incomplete data using importance sampling. *AISTATS*, Citeseer, 2005.
14. Stephan Grimm et al. Ontologies and the semantic we. In *handbook of Semantic Web Technology*, chapter 13 pages 507-579, 2011
15. Maria del Mar Roldan- García et al. Ontology-driven approach for KPI meta-modelling, selection and reasoning, *International Journal of Information Management*, 2021
16. <https://www.eurisportal.eu/>
17. <https://protege.stanford.edu/>
18. Hogan, A., Blomqvist, E., Cochez, M., d’Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4), 1-37.
19. Yu C., Wang F., Liu Y.H., An L., 2021, Research on knowledge graph alignment model based on deep learning, *Expert Systems with Applications*, 186, 115768.
20. Pearl J. *Causality*. Cambridge: Cambridge University Press; 2009.

21. Spirtes P, Glymour CN, Scheines R. Causation, Prediction, and Search. 2nd ed. Cambridge, MA: MIT Press; 2000.
22. Uusitalo L. Advantages and challenges of Bayesian networks in environmental modelling. *Ecol Model.* 2007;203:312,318.