STUDY PROTOCOL

REVISED Modelling temporal data in knowledge graphs: a systematic review protocol

[version 2; peer review: 1 approved, 2 approved with reservations]

Sepideh Hooshafza[1,2], Fabrizio Orlandi[2], Rachel Flynn[1], Louise McQuaid[1], Gaye Stephens[2], Laura O'Connor[1]

[1]Health Information and Quality Authority (HIQA), Cork, Ireland
[2]The SFI ADAPT Research Centre for AI-Driven Digital Content Technology, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

v2 **First published:** 10 Sep 2021, **4**:101
https://doi.org/10.12688/hrbopenres.13403.1
Latest published: 02 Aug 2022, **4**:101
https://doi.org/10.12688/hrbopenres.13403.2

Abstract

Background: The benefits of having high-quality healthcare data are well established. However, high-dimensionality and irregularity of healthcare data pose challenges in their management. Knowledge graphs have gained increasing popularity in many domains, as a method for representing data to overcome such challenges.
One important factor in representing data is "time". Data with time related attributes are considered, temporal data. Temporal data are frequently observed in healthcare and the management of rapidly changing patient data is an ongoing challenge. Traditionally, data models have focused on presenting static data and do not account for temporal data. Temporal data models ensure time consistency in data models and assist analysing the history of data and predicting the future trends in data. Knowledge graphs can include temporal data models and are therefore of interest to the field of healthcare data management.
As such, the herein aim is to outline a protocol for an inter-disciplinary systematic review of approaches, applications and challenges in modelling temporal data in knowledge graphs so that we can inform the application of knowledge graphs to healthcare data.
Method: The research questions is, what are the existing approaches in modelling temporal data in RDF based knowledge graphs. Two sub-questions on applications, and challenges will also be evaluated. ACM digital library, IEEE Xplore and Scopus will be searched for this review. The search will be limited to peer-reviewed literature referring to knowledge graphs based on Resource Description Framework (RDF). A narrative synthesis of the papers will be conducted.
Conclusion: The findings of this systematic review will be useful for data engineers to better represent data and perform analytics through temporal data. They can be applied in the context of healthcare data and the current challenges faced in managing rapidly changing patient data.

Open Peer Review

Approval Status ✓ ? ?

	1	2	3
version 2 (revision) 02 Aug 2022	✓ view		? view
version 1 10 Sep 2021	? view	? view	

1. **Bridget Kane**, Karlstad University, Karlstad, Sweden
2. **Paul Stacey**, Technological University Dublin, Dublin, Ireland
3. **Enrique Iglesias**, Leibniz University of Hannover, Hannover, Germany

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
Knowledge graph, temporal data, resource description framework

Corresponding author: Laura O'Connor (loconnor@hiqa.ie)

Author roles: Hooshafza S: Conceptualization, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Orlandi F**: Conceptualization, Methodology, Writing – Review & Editing; **Flynn R**: Conceptualization, Funding Acquisition, Methodology; **McQuaid L**: Conceptualization, Methodology, Supervision, Writing – Review & Editing; **Stephens G**: Conceptualization, Methodology, Supervision, Writing – Review & Editing; **O'Connor L**: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was conducted as part of the fulfilment of a PhD. The PhD scholar (SH) is funded by the Health Information and Quality Authority. Her supervisors (LMC and LOC) are funded by the Health Information Quality Authority and the Health Research Board.

How to cite this article: Hooshafza S, Orlandi F, Flynn R *et al.* **Modelling temporal data in knowledge graphs: a systematic review protocol [version 2; peer review: 1 approved, 2 approved with reservations]** HRB Open Research 2022, **4**:101 https://doi.org/10.12688/hrbopenres.13403.2

First published: 10 Sep 2021, **4**:101 https://doi.org/10.12688/hrbopenres.13403.1

REVISED **Amendments from Version 1**

Based on reviewers' comments on version 1, we made the following changes to our manuscript:

•Some additional texts have been added to the introduction to reflect more clarity on the rational for temporal data modelling in healthcare, and to explain existing approaches of temporal data modelling and their limitations.

•ScienceDirect database has been replaced with Scopus to cover more articles.

•Research question has been narrowed by adding RDF based knowledge graphs.

•Some additional rational has been provided for the selection of terms for the search strategy. Table 1 has been extended by adding PEO framework components. Search concepts and keywords have been described using PEO framework.

•The exact search queries have been placed for each databases.

•The following sentence has been added to the conclusion to broaden the area: "The findings of the systematic review will be of interest to organisations working in the fields of data science, data engineering, informatics and other similar areas."

Any further responses from the reviewers can be found at the end of the article

Introduction

The benefits of having high-quality, up-to-date, usable healthcare data are well established[1]. The healthcare data come from different sources such as hospitals, patient registries, clinics and are collected over time[2]. These sources generate large amounts of healthcare data such as patients' medical histories, physicians' notes, prescriptions, laboratory results and scan reports[3]. High-dimensionality, irregularity and sparsity of healthcare data pose challenges in their management, processing and usability[4]. Moreover, the volume of data generated in healthcare setting are increasing rapidly and makes it complicated for managing and analysing data[2]. As such, there is a need for effective methods for healthcare data storage and representation.

An important factor in storing data is "time"[5]. Time-varying data (also called temporal data) are data that have a time related attribute. Temporal data are created by including timestamps for the data values[6,7]. Timestamps in a data model are mostly used to indicate time points (valid time) in which the data values are valid and transaction time in which the data values are recorded[6,8,9]

Most healthcare data are temporal in nature[10]. Most of the clinical data such as patients' symptoms, laboratory test results, data from health monitoring devices, and various clinical measurements, are associated with a time in which the data is valid (valid time). Furthermore, capturing the time that the data are inserted into the database (transaction time) is required for legal purposes, treatment purposes or for justifying physicians' decisions[10]. As a simple example, suppose a patient has been hospitalised on February 18th at 8:00 am. This is the valid time for the hospitalisation time. This piece of information is inserted into hospital database on Feb 18 at 12:00 pm. This time is called the transaction time[10]. Capturing both transaction time and valid time is important to ensure that the decision making process is accurate and valuable and is beneficial in designing decision support systems[11].

Improved management of temporal healthcare data would benefit healthcare practitioners in information retrieval, healthcare decision making and support patient care[12]. Temporal data can assist in exploring temporal patterns in diseases and identifying disease progression. It would help in assessing patients' clinical history, finding possible causes of clinical events over time, and predicting future trends and events based on past and current clinical data[2]. Hence, the representation and query of temporal data has become a priority research area and efficient solutions are needed to model and store temporal data in healthcare settings[13,14].

Most of the solutions implemented so far for storing, and representing valid time and transaction time are based on relational databases. The developers of clinical database systems have some difficulty managing time values in relational databases for highly connected data. Firstly, modelling relationships and semantics between data items is not easy to implement in relational databases[15]. Secondly, for the large volume of data coming from heterogeneous sources, it is difficult to query a large number of joining tables[16]. As such, knowledge graphs (KGs) are potential solution to these problems.

In recent years, knowledge graphs (KGs) have been used in academic and industry as a method for managing and representing data[17,18]. They have attracted attention in several application areas including natural language processing, question answering machines, recommendation system[19].

KGs are defined as a semantic network comprising entities (nodes) and relationships (edges)[14]. There are two main types of KGs adhering to the Resource Description Framework (RDF) data model or the property graphs model[20]. For a number of reasons, the focus of this review is on RDF based knowledge graphs. RDF is a standard language for data representation and interchange on the Web[21]. RDF graphs are popular in practice and follow the World Wide Web Consortium (W3C) standards[22]. A community of practice and supporting tools have developed around the RDF and related semantic web standards. Many standards based public data models (called ontologies) are available to support and guide RDF KG enhancement and development.

Previous studies in the field of knowledge graphs focused on static data, however, methods to deal with and capture the variation and development of data over time, is of high importance and little is known about presenting temporal data in knowledge graphs[23,24]. Storing data by considering time varying knowledge, ensures time consistency in a data model, improve performance of KG models, and assist analysing the history of data and predicting the future trends in data as well[5,8].

The herein aim is to outline a protocol for a systematic review to explore existing approaches, applications and challenges in modelling temporal data in knowledge graphs. The results of the systematic review will inform data engineers and others of the feasibility and challenges involved when modelling temporal data. In healthcare, the findings of this study will assist in modelling patient data, data from health monitoring devices and data collected within services over time. There are different international health information modelling standards including OpenEHR, HL7 which are used for modelling information at high level and for semantic interoperability purposes. In this review our focus is on the modelling in the data level, where we can model data to store them in the database and query them[25,26]. International health information modelling standards can be used to get a high level overview of the concept with its included information items and can be used for designing different concepts in knowledge graphs but they are not replaceable.

The management of rapidly changing patient data and types of data sets is an ongoing challenge. This study focuses on the approaches and challenges in KG modelling to support this management[2,12].

Protocol

Research methodology

This systematic review is based on the guidelines and procedures for systematic reviews within the software engineering domain[27,28].

The procedure that will be undertaken in this study is as follows:

1. Formulating the research questions;
2. Selecting information sources (digital libraries) on which to perform search;
3. Defining search concepts and keywords;
4. Application of search terms on databases;
5. Considering inclusion and exclusion criteria for selection of studies;
6. Quality appraisal of the included studies;
7. Synthesis of data.

Research question

Major research question: What are the existing approaches in modelling temporal data in RDF based knowledge graphs?

Sub-questions:

1. What are the existing applications of temporal data models in RDF based knowledge graphs?
2. What are the existing challenges with modelling temporal data in RDF based knowledge graphs?

Information sources

Searches will be carried out on the following databases: ACM Digital Library[29], IEEE Xplore Digital Library[30], and Scopus[31]. The bibliographies of the included full-text articles will be searched for relevant articles. Searching of forward citations will also be conducted to identify other potential material for inclusion.

Search strategy

The search strategy was developed using the PEO (Population (Context), Exposure, and Outcome) framework as follows[32], Population (Context): knowledge graphs, Exposure: Time, and Outcome: Applied model. The search concepts and keywords based on PEO framework are set out in Table 1.

The search query to be used for ACM digital library and Scopus, is:

*"((“Knowledge graph” OR rdf OR “resource description framework”) AND (Temporal OR dynamic OR evolution OR time) AND (*present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied))"*

Since IEEE Xplore Digital Library does not accept using parenthesis in the advanced search, the search query to be used is:

“Knowledge graph” OR rdf OR “resource description framework”

AND

Temporal OR dynamic OR evolution OR time

AND

Table 1. Search terms for a systematic review on modelling temporal data in knowledge graphs.

PEO framework	Concepts	Search terms
Population (context)	Knowledge graph	"Knowledge graph" OR RDF OR "resource description framework"
Exposure	Temporal	Temporal OR dynamic OR evolution OR time
Outcome	Applied model	*present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied

**present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied*

Criteria for inclusion

No limits will be applied to articles for inclusion in terms of publication date or language.

Articles will be included if they:

- Refer to approaches in modelling temporal data in KGs

  OR

- Discuss applications of temporal data modelling in KGs

  OR

- Address challenges of temporal data modelling in KGs

  AND

- Refer to knowledge graphs based on Resource Description Framework (RDF)

Studies will be excluded if they refer to knowledge graphs based on frameworks other than Resource Description Framework (RDF).

Types of study to be included

Peer-reviewed literature will be selected to be reviewed in this study. Given the nature of the topic under review, it is anticipated that the studies will mostly fall into the category of original research articles.

Software

Covidence by Veritas Health Innovation Ltd, a web-based software platform for systematic review management, will be used for screening articles[33]. EndNote X8.2 by PDF Tron™ Systems Inc. will be used to manage the bibliography[34]. Microsoft Excel will be used to manage the extracted data.

Screening

All retrieved articles from the selected information sources will be imported into Covidence. Duplicate references will be removed. Two reviewers will independently screen the titles and abstracts against the inclusion/exclusion criteria. Any disagreements on inclusion/exclusion will be firstly resolved by discussion. Any disagreements not resolved by discussion will be resolved by a third author. Forward citation and hand-searching of bibliographies of included studies will be performed and any relevant studies identified will be included. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement will be used to report the search and study selection process[35].

Quality appraisal

A quality appraisal tool will be used to inform weighting of discussion based on the quality of the included articles. The quality appraisal checklist proposed in the Guidelines for performing systematic literature reviews in software engineering will be used for this purpose[27]. Two reviewers will independently appraise the quality of selected articles. If agreement cannot be reached, a third researcher will assess the studies to come to a consensus. Articles will not be excluded based on their quality.

Data extraction

A data extraction table will be developed in Covidence to structure and categorise the findings (See extended data). The data to be extracted in the table includes study ID, study title, author(s), publication type, year of publication, journal/conference title, setting, modelling approaches, applications, and challenges in modelling temporal data in knowledge graphs. Once the table is completed for all the final included articles, the table will be exported to Excel and data synthesis will be conducted.

The data extraction table will be piloted on three articles by two researchers to ensure appropriateness of the included data extraction fields against the data provided in articles and mutual understanding of the fields. The data extraction table will be updated at this point, if required.

Data synthesis

The information will be manually extracted from each included article. Articles will be read in full by one researcher and the data extracted directly into the data extraction table. A second researcher will independently complete data extraction for 10% of the identified articles for quality assurance purposes. A narrative synthesis will be performed to analyse the articles. To facilitate the visualization of the information, the synthesis of the extracted data will be presented in different forms including tables, graphs and other artefacts.

Dissemination of information

The systematic review will be submitted to an academic journal on completion. Conference abstracts arising out of the systematic review will also be submitted to appropriate conferences for presentation.

Strengths and limitations

To the best of the authors' knowledge, this review will be the first to systematically describe temporal data modelling in knowledge graphs. In addition, the methodological approach allows for a comprehensive exploration of modelling approaches, applications, and challenges of temporal data modelling in knowledge graphs. A further strength of this review is that the search is not limited to the field of healthcare information. It has been designed so that we can gain learning from across the disciplines and use that to inform practice in health information management.

In terms of limitations, due to the multiplicity of concepts and keywords used in the literature, there is a risk that some relevant studies may not be retrieved. This risk has been reduced by evaluating a range of studies in preliminary searches to ensure that equivalent words and phrases are included in the search terms. Furthermore, the inclusion of hand-searching of bibliographies and forward citation searching is designed to, in part, overcome this limitation.

Conclusions

The purpose of conducting this review is to identify existing approaches and applications of modelling temporal data in knowledge graphs and to identify challenges of modelling temporal data in knowledge graphs. The findings of the systematic review will be of interest to organisations working in the fields of data science, data engineering, informatics and other similar areas. They can also inform quality improvement initiatives for health information system service providers and help generate new ideas in temporal healthcare data modelling and develop data analytics solution based on temporal healthcare data. This will be beneficial in addressing the current challenges faced in managing rapidly changing patient data.

Study status

Searching the information sources using the search terms outlined in Table 1 has commenced.

Data availability

Underlying data

No data are associated with this article.

Extended data

Figshare: Data Extraction Table_Systematic Review_SH 2021.docx. https://doi.org/10.6084/m9.figshare.16528308[36]

This protocol contains the following extended data: Data Extraction Table_Systematic Review_SH 2021.docx. (Data extraction table)

Reporting guidelines

Figshare: PRISMA-P checklist for "Modelling temporal data in knowledge graphs: a systematic review protocol". https://doi.org/10.6084/m9.figshare.16499031[37]

Acknowledgments

The assistance provided by the researchers from the school of computer science and statistics of Trinity College Dublin is greatly appreciated.

Supplementary material

This study does not contain any supplementary material.

References

1. Catalyst N: **Healthcare Big Data and the Promise of Value-Based Care**. 2018; Accessed on: 10 August 2021.
Reference Source
2. Poh N, Tirunagari S, Windridge D: **Challenges in designing an online healthcare platform for personalised patient analytics**. *2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD);*. 2014 9-12 Dec 2014.
Publisher Full Text
3. Archenaa J, Anita EAM: **Interactive Big Data Management in Healthcare Using Spark**. *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC – 16')*. Cham: Springer International Publishing. 2016; 265–272.
Publisher Full Text
4. Lee C, Luo Z, Ngiam KY, *et al.*: **Big Healthcare Data Analytics: Challenges and Applications**. In: Khan SU, Zomaya AY, Abbas A,editors. *Handbook of Large-Scale Distributed Computing in Smart Healthcare*. Cham: Springer International Publishing; 2017; 11–41.
Publisher Full Text
5. Xiaoyi Z, Jing Q, Qing W: **Research on the solution to redundancy of temporal data.** *2010 International Conference on Computer Application and System Modeling (ICCASM 2010).* 2010; 22–24.
Publisher Full Text
6. Kumar S, Rishi R: **A relative analysis of modern temporal data models.** *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).* 2016; 16–18.
Reference Source
7. Moere AV: **Time-Varying Data Visualization Using Information Flocking Boids.** *IEEE Symposium on Information Visualization.* IEEE; 2004; 97–104.
Publisher Full Text
8. Ni R, Ma Z, Yu K, *et al.*: **Specific Time Embedding for Temporal Knowledge Graph Completion.** *2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC).* 2020; 26–28.
Publisher Full Text
9. Wang Z, Li X: **Hybrid-TE: Hybrid Translation-Based Temporal Knowledge Graph Embedding.** *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI).* 2019; 4–6.
Publisher Full Text
10. Stantic B, Terenziani P, Sattar A: **Coping efficiently with now-relative medical data.** *AMIA Annu Symp Proc.* 2008; **2008**: 722–6.
PubMed Abstract | **Free Full Text**
11. Johnston T: **Bitemporal data: theory and practice**. Newnes; 2014.
Reference Source
12. Combi C, Cucchi G, Pinciroli F: **Applying object-oriented technologies in modeling and querying temporally oriented clinical databases dealing with temporal granularity and indeterminacy.** *IEEE Trans Inf Technol Biomed.* 1997; **1**(2): 100–27.
PubMed Abstract | **Publisher Full Text**
13. Li H: **A New query method for the temporal RDF Model RDFMT Based on SPARQL**. *2021 2nd International Conference on Artificial Intelligence and Information Systems*; Chongqing, China: Association for Computing Machinery; 2021; 1–6.
Publisher Full Text
14. Wang J, Zhang W, Chen X, *et al.*: **3DRTE: 3D Rotation Embedding in Temporal Knowledge Graph.** *IEEE Access.* 2020; **8**: 207515–207523.
Publisher Full Text
15. Adlassnig KP, Combi C, Das AK, *et al.*: **Temporal representation and reasoning in medicine: Research directions and challenges.** *Artif Intell Med.* 2006; **38**(2): 101–13.
PubMed Abstract | **Publisher Full Text**

16. Medhi S, Baruah H: **Relational database and graph database: A comparative analysis.** *Journal of Process Management New Technologies.* 2017; **5**: 1–9.
Publisher Full Text

17. Zou X: **A Survey on Application of Knowledge Graph.** *J Phys Conf Ser.* 2020; **1487**: 012016.
Publisher Full Text

18. Yan J, Wang C, Cheng W, *et al.*: **A retrospective of knowledge graphs.** *Front Comput Sci.* 2018; **12**(1): 55–74.
Publisher Full Text

19. Chen Z, Wang Y, Zhao B, *et al.*: **Knowledge Graph Completion: A Review.** *IEEE Access.* 2020; **8**: 192435–192456.
Publisher Full Text

20. Angles R, Thakkar H, Tomaszuk D: **Mapping RDF Databases to Property Graph Databases.** *IEEE Access.* 2020; **8**: 86091–86110.
Publisher Full Text

21. Daniele DA, Emanuele Della V, Jean-Paul C, *et al.*: **RSP-QL Semantics: A Unifying Query Model to Explain Heterogeneity of RDF Stream Processing Systems.** *Int J Semant Web Inf Syst.* 2014; **10**(4): 17–44.
Reference Source

22. Kim YH, Kim BG, Lim HC: **The index organizations for RDF and RDF schema.** *2006 8th International Conference Advanced Communication Technology.* 2006; 20–22.
Publisher Full Text

23. Ahn YS, Jeong OR: **Time-Aware PolarisX: Auto-Growing Knowledge Graph.** *Comput Mater Contin.* 2021; **67**(3): 2695–708.
Publisher Full Text

24. Xu C, Nayyeri M, Alkhoury F, *et al.*: **Temporal Knowledge Graph completion based on time series Gaussian embedding.** *International Semantic Web Conference.* 2020; **12506**: 654–671.
Publisher Full Text

25. **OpenEHR**.
Reference Source

26. **HL7**.
Reference Source

27. Kitchenham B: **Procedures for Performing Systematic Reviews.** Keele, UK, Keele Univ. 2004; 33.

28. Dyba T, Dingsoyr T, Hanssen GK, *et al.*: **Applying Systematic Reviews to Diverse Study Types: An Experience Report**. *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*. 2007.
Publisher Full Text

29. **ACM Digital Library**. Accessed on: 31 July 2021.
Reference Source

30. **IEEE Xplore digital library**. Accessed on: 31 July 2021.
Reference Source

31. **Scopus**.
Reference Source

32. **How to Conduct a Literature Review (Health Sciences and Beyond)**.
Reference Source

33. **Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia**. Accessed on: 31 July 2021.
Reference Source

34. **EndNote™ X8- The best reference management tool**. Accessed on: 31 July 2021.
Reference Source

35. **Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)**. Accessed on: 31 July 2021.
Reference Source

36. Hooshafza S, Orlandi F, Flynn R, *et al.*: **Data Extraction Table_Systematic Review_SH 2021.docx**. 2021; Accessed on: 27 August 2021.
http://www.doi.org/10.6084/m9.figshare.16528308

37. Hooshafza S, Orlandi F, Flynn R, *et al.*: **PRISMA-P_Checklist_Systematic_ Review_SH_ 2021_docx**. 2021; Accessed on: 27 August 2021.
http://www.doi.org/10.6084/m9.figshare.16499031

Open Peer Review

Current Peer Review Status: ✔ ? ?

Version 2

Reviewer Report 01 March 2024

https://doi.org/10.21956/hrbopenres.14859.r38059

?

Enrique Iglesias
Leibniz University of Hannover, Hannover, Germany

General Comments

Summary

This paper proposes a study that analyses existing approaches, applications, and challenges in modeling temporal data. The purpose of this study is to determine which existing approaches and works that will help inform engineers on the challenges of modeling temporal data. Additionally, this work seeks to inform healthcare workers about the benefits of using temporal data in knowledge graphs.

This paper describes the study the authors conducted to find existing works illustrating the benefits and challenges of modeling temporal data. The authors used multiple databases like ACM Digital Library, IEEE Xplore Digital Library, and Scopus to extract relevant works. Additionally, the authors search the references of the relevant papers for additional papers that will enrich the study.

This paper has two main negative points. First, some sentences in the introduction need to be rewritten. Some sentences could be more understandable. For example, “A community of practice and supporting tools have developed around the RDF and related semantic web standards.” This sentence should be “A community of practice and supporting tools have been developed around RDF and related semantic web standards.” This sentence would be good to provide examples of these tools and highlight which aspects of RDF are being worked on. Other sentences need to be divided into shorter sentences. For example, the sentence “Previous studies in the field of knowledge graphs focused on static data, however, methods to deal with and capture the variation and development of data over time, is of high importance and little is known about presenting temporal data in knowledge graphs” is way too long and it needs to be divided into at least two smaller sentences if not three sentences. Some concepts are not used correctly, like “improve the performance of KG models,” a knowledge graph doesn’t have performance. Therefore, are the models in question the creation process or the querying? They need to be explained better. The last penultimate paragraph of the introduction needs to be reordered. The paragraph tries to summarize the aim of the paper, but it jumps back in forth between the purpose of the paper and the benefits of using temporal data in a health case setting. Finally, the last paragraph doesn’t close the introduction well. I would recommend restructuring the

penultimate paragraph for the closing of the paragraph. The second main negative issue is that even though the methodology behind the study is well-defined, the results of the study aren't actually presented. Therefore, the actual impact of the study can only be assumed but not confirmed.
This paper presents a well-defined study to find existing works illustrating the benefits and challenges of modeling temporal data. Unfortunately, the paper needs rewriting in some parts, as well as fixing some typos.

Section Comments

Section 1: Introduction

- Positive: This section does a good job of introducing the reader to the purpose of this work, as well as the needed background knowledge to understand this work.
- Negative: Some sentences in the introduction need to be rewritten. Some sentences could be more understandable. For example, "A community of practice and supporting tools have developed around the RDF and related semantic web standards." This sentence should be "A community of practice and supporting tools have been developed around RDF and related semantic web standards." This sentence would be good to provide examples of these tools and highlight which aspects of RDF are being worked on. Other sentences need to be divided into shorter sentences. For example, the sentence "Previous studies in the field of knowledge graphs focused on static data, however, methods to deal with and capture the variation and development of data over time, is of high importance and little is known about presenting temporal data in knowledge graphs" is way too long and it needs to be divided into at least two smaller sentences if not three sentences. Some concepts are not used correctly, like "improve the performance of KG models," a knowledge graph doesn't have performance. Therefore, are the models in question the creation process or the querying? They need to be explained better. In some cases where items are being listed do not have an "and" in the last item. For example, "natural language processing, question answering machines, recommendation system" doesn't have an "and" between question answering machines and recommendation system. This can also be seen in "OpenEHR, HL7." The acronym for knowledge graph is introduced twice (one right after the other) and is only used two more times afterward and never again in the rest of the paper. The last penultimate paragraph of the introduction needs to be reordered. The paragraph tries to summarize the aim of the paper, but it jumps back in forth between the purpose of the paper and the benefits of using temporal data in a health case setting. Finally, the last paragraph doesn't close the introduction well. I would recommend restructuring the penultimate paragraph for the closing of the paragraph.

Section 2: Protocol

- Positive: This section defines all the characteristics of the study that will conducted, including the databases used to search for the relevant papers, queries used for each particular database, and the criteria in which the papers are considered for the study.
- Negative: This section presents a couple of sentences I recommend changing to make them cleaner. "The bibliographies of the included full-text articles will be searched for relevant articles" would be better written as "The bibliographies of the included full-text articles will be searched for additional relevant articles." In the sentence "The search strategy was developed using the PEO (Population (Context), Exposure, and Outcome) framework as follows," it's strange to have parentheses within parentheses. Finally, the sentence "The data to be extracted in the table" should be "The data to be extracted into the table."

Section 3: Conclusion

- Positive: This section closes the paper well. It is a good summary of what is presented in the paper and its possible impact.
- Negative: The main issue with this section is that the paper doesn't have any results to support the impact of the study. Therefore, any conclusion drawn from the paper can only be considered a prediction. Additionally, there is a small typo in "organisations"; it should be "organizations."

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Knowledge graph creation, Data managment

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 15 September 2022

https://doi.org/10.21956/hrbopenres.14859.r32619

Bridget Kane
Karlstad University, Karlstad, Sweden

I have reviewed the changes. I am satisfied with the amendments made by the authors and I formally approve this publication.

Is the rationale for, and objectives of, the study clearly described?

Not applicable

Is the study design appropriate for the research question?

Not applicable

Are sufficient details of the methods provided to allow replication by others?
Not applicable

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health Information Systems, eHealth, human-computer interaction (HCI) in healthcare and computer-supported collaborative work (CSCW) in healthcare.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 14 December 2021

https://doi.org/10.21956/hrbopenres.14603.r30875

?

Paul Stacey
[1] School of Informatics and Engineering, Technological University Dublin, Dublin, Ireland
[2] School of Informatics and Engineering, Technological University Dublin, Dublin, Ireland

This article proposes an interesting protocol for a systematic review to investigate current approaches to the modelling of temporal data in knowledge graphs. The protocol is based on pre-existing guidelines, but is unique in it's application to establishing a review of existing approaches in modelling temporal data in knowledge graphs and some of the underlying methods employed within the protocol.

The purpose of the review is to ultimately investigate the use of knowledge graphs in representing, storing and managing complex healthcare data. However, this will be the focus of future work. This article describes the approach to be taken to determine the current body of work in using knowledge graphs and specifically RDF to handle temporal models or the application of the latter to the former. From that perspective the article is of interest to the health informatics research community.

The article is well written and structured well, allowing the reader to easily navigate the material and follow the argument and the protocol proposed. However, there are a few areas within the

article that should be reviewed to aide clarity for the reader. These are outlined below:

- The introduction makes a big jump from stating the need for effective methods for healthcare data storage and representation and in particular dealing with temporal data without mentioning some key approaches already employed. The argument for the "need" isn't well established in the text at this point.

- Knowledge Graphs are proposed as a solution somewhat abruptly. It would be helpful for the reader to include some additional background here on the challenges that current approaches suffer when representing temporal data and why KGs could be the solution.

- The term healthcare data is very broad, and there are examples of using knowledge graphs within the literature for some specific use cases and sub sets of healthcare data. But is the focus on Patient data/information? Or perhaps it is all healthcare data and a big data approach? I suspect it is all data given the reference material to big data source material, however this is not overly clear in the text.

- Some additional rationale should be given for the selection of terms for the search strategy.

- The research question states: "What are the existing approaches in modelling temporal data in knowledge graphs?" - but the study is only focused on articles mentioning RDF. However, RDF is not the only way to represent knowledge graphs as is acknowledged later in the text. Therefore the research question should be narrowed or the study design broadened to all knowledge graphs or at least a rationale for only focusing on RDF should be provided for the reader.

- The statement "However, little is known about the use or suitability for use with healthcare data." is a very broad statement and the literature would show that there is a reasonable, if not limited body of knowledge available in the literature investigating KGs applicability to healthcare data . But, again, perhaps it is the particular type of healthcare data that the authors are referring to, this is unclear and should be reviewed.

- The authors make the argument that temporal data is important to be represented, but don't link this to healthcare data approaches already employed within the health informatics community to represent temporal data, for example: Data Types Information Model (openehr.org). What are the limitations of these approaches and why are knowledge graphs important etc? Or perhaps the author seeks to apply these approaches to knowledge graphs and the reification approaches used such as RDF; this is unclear from the reader's perspective.

- The statement "The results of the systematic review will inform data engineers when modelling temporal data". This statement is somewhat reductive and should be broadened to something like "informaticians and the data science community".

- Would the authors consider using Scopus instead of ScienceDirect as one of the information sources? Scopus would give a broader coverage from additional publishers beyond that of ScienceDirect, while also covering the metadata from ScienceDirect.

- Reference 28 and 29 appear to link to the same Figure?
- Some minor grammatical errors and typos exist that should be reviewed, for example: In the Abstract space required between "time”.Data".

Is the rationale for, and objectives of, the study clearly described?
Partly

Is the study design appropriate for the research question?
Partly

Are sufficient details of the methods provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Information modelling, semantic systems, knowledge based systems, internet of things.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 28 Jul 2022

Sepideh Hooshafza

This article proposes an interesting protocol for a systematic review to investigate current approaches to the modelling of temporal data in knowledge graphs. The protocol is based on pre-existing guidelines, but is unique in it's application to establishing a review of existing approaches in modelling temporal data in knowledge graphs and some of the underlying methods employed within the protocol.

The purpose of the review is to ultimately investigate the use of knowledge graphs in representing, storing and managing complex healthcare data. However, this will be the focus of future work. This article describes the approach to be taken to determine the current body of work in using knowledge graphs and specifically RDF to handle temporal models or the application of the latter to the former. From that perspective the article is of interest to the health informatics research community.

The article is well written and structured well, allowing the reader to easily navigate the material and follow the argument and the protocol proposed. However, there are a few areas within the article that should be reviewed to aide clarity for the reader. These are outlined below:

1. The introduction makes a big jump from stating the need for effective methods for healthcare data storage and representation and in particular dealing with temporal data without mentioning some key approaches already employed. The argument for the "need" isn't well established in the text at this point.

Thank you for your comment. We have now included additional text to clarify the need for temporal data modelling specifically in healthcare with a brief introduction of temporal data in healthcare, as below:

"Most healthcare data are temporal in nature.[10] Most of the clinical data such as patients' symptoms, laboratory test results, data from health monitoring devices, and various clinical measurements, are associated with a time in which the data is valid (valid time). Furthermore, capturing the time that the data are inserted into the database (transaction time) is required for legal purposes, treatment purposes or for justifying physicians' decisions.[10] As a simple example, suppose a patient has been hospitalised on February 18th at 8:00 am. This is the valid time for the hospitalisation time. This piece of information is inserted into hospital database on Feb 18 at 12:00 pm. This time is called the transaction time.[10] Capturing both transaction time and valid time is important to ensure that the decision making process is accurate and valuable and is beneficial in designing decision support systems.[11]
Improved management of temporal healthcare data would benefit healthcare practitioners in information retrieval, healthcare decision making and support patient care.[12] Temporal data can assist in exploring temporal patterns in diseases and identifying disease progression. It would help in assessing patients' clinical history, finding possible causes of clinical events over time, and predicting future trends and events based on past and current clinical data.[2] Hence, the representation and query of temporal data has become a priority research area and efficient solutions are needed to model and store temporal data in healthcare settings.[13,14] "

And

"Temporal data models are a way to ensure time consistency in data models and assist analysing the history of data and predicting the future trends in data. The management of rapidly changing patient data and types of data sets is an ongoing challenge. This study focuses on the approaches and challenges in KG modelling to support this management.[3, 4]"

2. Knowledge Graphs are proposed as a solution somewhat abruptly. It would be helpful for the reader to include some additional background here on the challenges that current approaches suffer when representing temporal data and why KGs could be the solution.

Thank you for your comments. We have included the following text to address this comment:

"Most of the solutions implemented so far for storing, and representing valid time and transaction time are based on relational databases. The developers of clinical database systems have some difficulty managing time values in relational databases for highly connected data. Firstly, modelling relationships and semantics between data items is not easy to implement in

relational databases.[15] Secondly, for the large volume of data coming from heterogeneous sources, it is difficult to query a large number of joining tables.[16] As such, knowledge graphs (KGs) are potential solution to these problems."

3. The term healthcare data is very broad, and there are examples of using knowledge graphs within the literature for some specific use cases and sub sets of healthcare data. But is the focus on Patient data/information? Or perhaps it is all healthcare data and a big data approach? I suspect it is all data given the reference material to big data source material, however this is not overly clear in the text.

We apologise that this was not clear. We mean all healthcare data including patient data, data from health monitoring devices, data collected within services. We have included some examples of healthcare data in the introduction to clarify this further:

"Most healthcare data are temporal in nature.[10] Most of the clinical data such as patients' symptoms, laboratory test results, data from health monitoring devices, and various clinical measurements, are associated with a time in which the data is valid (valid time). Furthermore, capturing the time that the data are inserted into the database (transaction time) is required for legal purposes, treatment purposes or for justifying physicians' decisions.[10] As a simple example, suppose a patient has been hospitalised on February 18th at 8:00 am. This is the valid time for the hospitalisation time. This piece of information is inserted into hospital database on Feb 18 at 12:00 pm. This time is called the transaction time.[10] Capturing both transaction time and valid time is important to ensure that the decision making process is accurate and valuable and is beneficial in designing decision support systems.[11]

4. Some additional rationale should be given for the selection of terms for the search strategy.

We have included additional text in the search strategy section as follows:

"The search strategy was developed using the PEO (Population (Context), Exposure, and Outcome) framework as follows, Population (Context): knowledge graph, Exposure: Time, and Outcome: Applied model. The search concepts and keywords based on PEO framework are set out in Table 1 .

Table 1- Search terms for a systematic review on modelling temporal data in knowledge graphs

PEO framework	Concepts	Search terms
Population (Context)	knowledge graph	"Knowledge graph" OR RDF OR "resource description framework"
Exposure	Time	Temporal OR dynamic OR evolution OR time

Outcome
Applied model
*present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied

5. The research question states: "What are the existing approaches in modelling temporal data in knowledge graphs?" - but the study is only focused on articles mentioning RDF. However, RDF is not the only way to represent knowledge graphs as is acknowledged later in the text. Therefore the research question should be narrowed or the study design broadened to all knowledge graphs or at least a rationale for only focusing on RDF should be provided for the reader.

Thank you for your comment, we have updated the research question as follows:

"What are the existing approaches in modelling temporal data in RDF based knowledge graphs?"

There a number of reasons why the focus was on RDFs. We have also included text that highlights the rationale for using RDF graphs in the introduction to justify the focus of the review.

"For a number of reasons, the focus of this review is on RDF based knowledge graphs. RDF is a standard language for data representation and interchange on the Web.[21] RDF graphs are popular in practice and follow the World Wide Web Consortium (W3C) standards.[22] A community of practice and supporting tools have developed around the RDF and related semantic web standards. Many standards based public data models (called ontologies) are available to support and guide RDF KG enhancement and development."

6. The statement "However, little is known about the use or suitability for use with healthcare data." is a very broad statement and the literature would show that there is a reasonable, if not limited body of knowledge available in the literature investigating KGs applicability to healthcare data. But, again, perhaps it is the particular type of healthcare data that the authors are referring to, this is unclear and should be reviewed.

We agree with the reviewer and apologise that we were not tighter with the language. What was meant was limited data in relation to temporal healthcare data specifically. We have removed this statement and replaced it with:

"*Little is known about presenting temporal data in knowledge graphs*."

7. The authors make the argument that temporal data is important to be represented, but don't link this to healthcare data approaches already employed within the health informatics community to represent temporal data, for example: Data Types Information Model (openehr.org). What are the limitations of these approaches and why are knowledge graphs important etc? Or perhaps the author seeks to apply these approaches to knowledge graphs and the reification approaches used such as RDF;

this is unclear from the reader's perspective.

Thank you for your comment, we have included additional text in the introduction to address this. The additional text is as follows:

"There are different international health information modelling standards including OpenEHR, HL7 which are used for modelling information at high level and for semantic interoperability purposes. In this review our focus is on the modelling in the data level, where we can model data to store them in the database and query them. International health information modelling standards can be used to get a high level overview of the concept with its included information items and can be used for designing different concepts in knowledge graphs but they are not replaceable."

8. The statement "The results of the systematic review will inform data engineers when modelling temporal data". This statement is somewhat reductive and should be broadened to something like "informaticians and the data science community".

Thank you for your comment, we agree and we have now included the informatics area in the conclusion section as well, as follows:

"The findings of the systematic review will be of interest to organisations working in the fields of data science, data engineering, informatics and other similar areas."

9. Would the authors consider using Scopus instead of ScienceDirect as one of the information sources? Scopus would give a broader coverage from additional publishers beyond that of ScienceDirect, while also covering the metadata from ScienceDirect.

Thank you very much for your suggestion. We have reviewed the content of both databases and we agree with changing ScienceDirect to Scopus since it is more comprehensive. We have updated the protocol as follow:

"Searches will be carried out on the following databases: ACM Digital Library [(27)], IEEE Xplore Digital Library [(28)], and Scopus [(29)]."

10. Reference 28 and 29 appear to link to the same Figure?

Thank you for highlighting this, we apologise for this error. We have corrected this error.

11. Some minor grammatical errors and typos exist that should be reviewed, for example: In the Abstract space required between "time". Data".

Thank you for highlighting this. We have corrected this error. We have also reviewed the paper in full for grammatical errors and typos.

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 November 2021

https://doi.org/10.21956/hrbopenres.14603.r30297

?

Bridget Kane
[1] Karlstad University, Karlstad, Sweden
[2] Karlstad University, Karlstad, Sweden

This is an interesting protocol to undertake a systematic review on modelling temporal data in knowledge graphs.

Generally, the study is well described. However, it would be better if more detail could be given about the Resource Development Framework (RDF), its importance and the reasons why it is a critical inclusion criterion. Please also expand on why studies based on other frameworks will be excluded.

Other points that should be addressed:

- Information sources: Is there a reason why PubMed is not included?
- Search strategy: Isn't the idea of publishing a protocol is that the exact protocol is provided and can be replicated? Yet, the search strategy given here is called 'an example'.
- The search strategy should be definitive and clearly articulated.
- Rather than assume that data from health monitoring devices are included, it would be good to be more clear, i.e. explicitly state that knowledge graphs of such data are included.

Is the rationale for, and objectives of, the study clearly described?
Partly

Is the study design appropriate for the research question?
Yes

Are sufficient details of the methods provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health Information Systems, eHealth, human-computer interaction (HCI) in

healthcare and computer-supported collaborative work (CSCW) in healthcare.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 28 Jul 2022

Sepideh Hooshafza

This is an interesting protocol to undertake a systematic review on modelling temporal data in knowledge graphs.

1. Generally, the study is well described. However, it would be better if more detail could be given about the Resource Development Framework (RDF), its importance and the reasons why it is a critical inclusion criterion. Please also expand on why studies based on other frameworks will be excluded.

Thank you for your comment. There are a number of reasons why the focus was on RDFs. We have included the following text in the introduction that highlights the rationale for using RDF graphs to justify the focus of the paper.

"For a number of reasons, the focus of this review is on RDF based knowledge graphs. RDF is a standard language for data representation and interchange on the Web.[21] RDF graphs are popular in practice and follow the World Wide Web Consortium (W3C) standards.[22] A community of practice and supporting tools have developed around the RDF and related semantic web standards. Many standards based public data models (called ontologies) are available to support and guide RDF KG enhancement and development."

Other points that should be addressed:

2. Information sources: Is there a reason why PubMed is not included?

PubMed is a repository focusing on life sciences and biomedical topics. We decided to include computer science databases and explore the area in general and not specifically in healthcare settings. The aim was to give a general view of the approaches, applications and challenges of temporal data modelling in knowledge graphs. We have as such not included PubMed despite the reviewer's observation.

3. Search strategy: Isn't the idea of publishing a protocol is that the exact protocol is provided and can be replicated? Yet, the search strategy given here is called 'an example'.

Thank you for your comment. We were following the PRISMA guidelines on providing an example of a search for one database. However, the PRISMA guidelines have been updated and now suggest including all searches in the protocol. As such we have now included the exact search strategy in the protocol and we have removed the word "example" from the text.

"The search query to be used for ACM digital library, and Scopus, is:

*((Temporal OR dynamic OR evolution OR time) AND ("Knowledge graph" OR rdf OR "resource description framework") AND (*present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied))*

Since IEEE Xplore Digital Library does not accept using parenthesis in the advanced search, the search query to be used is:
"Knowledge graph" OR rdf OR "resource description framework"
AND
Temporal OR dynamic OR evolution OR time
AND
**present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied "*

4. The search strategy should be definitive and clearly articulated.

As above, the exact search strategy has now been written in the protocol.

"The search query to be used for ACM digital library, and Scopus, is:

*((Temporal OR dynamic OR evolution OR time) AND ("Knowledge graph" OR rdf OR "resource description framework") AND (*present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied)) "*

Since IEEE Xplore Digital Library does not accept using parenthesis in the advanced search, the search query to be used is:
"Knowledge graph" OR rdf OR "resource description framework"
AND
Temporal OR dynamic OR evolution OR time
AND
**present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied "*

5. Rather than assume that data from health monitoring devices are included, it would be good to be clearer, i.e. explicitly state that knowledge graphs of such data are included.

We did not want to specify health monitoring devices because that level of specificity may have skewed the study where we are attempting to cover all health data. However, we acknowledge take your point and have included health care monitoring devices in the list of examples so as to be more explicit, as follows:

"Most healthcare data are temporal in nature.[10] Most of the clinical data like patients' symptoms, laboratory test results, data from health monitoring devices, and various clinical measurements, are associated with a time in which the data is valid (valid time)."

Competing Interests: No competing interests were disclosed.