The Persistence of Temporality: Representation of Time in Cultural Heritage Knowledge Graphs

Oleksandra Bruns[1,2]

[1]FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[2]Karlsruhe Institute of Technology (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany

Abstract
Modeling temporal information in cultural heritage knowledge graphs is essential to understand history, society and the world. However, due to the binary nature of RDF predicates, representing time in RDF is a challenge. This work addresses the challenge and explores various approaches for integrating time into knowledge graphs. In particular, the work aims at answering key research questions, such as best practices for temporal representation in RDF, crucial requirements in the cultural heritage context, and evaluating RDF extensions for modeling time. Additionally, it is investigated how RDF can reason over temporal relationships and how the new insights can be used to develop a lightweight and intuitive approach for modeling time in RDF based on domain requirements. Fundamentally, the research argues for embracing the persistence of temporality in KGs, as it is crucial for understanding our changing heritage, recognising the influence of the past on the present, and shaping a knowledgeable future.

Keywords
Temporal Data, Knowledge Graphs, Cultural Heritage

1. Introduction

Cultural heritage (CH) knowledge graphs (KGs) are an essential tool for representing complex relationships between artifacts, practices, events and traditions, enabling a better view on our shared heritage. A structured semantic representation of CH data improves interoperability, facilitating global access and reuse. Understanding historical events across different time layers enables comprehending their influence on each other and the present world, as well as fostering cultural identity. Hence, temporal context is crucial in the CH domain. For example, KGs can provide innovative ways of exploring diverse archival records from various historical periods by semantically interlinking them. However, this requires accurate representation, reasoning, and querying of temporal data and knowledge contained in such records.

The Resource Description Framework (RDF) is a widely adopted representation language and exchange format used to encode data on the Web, and a fundamental building block for creating knowledge graphs. However, the standard RDF model has limited capabilities in representing temporal aspects of data. Several approaches have been proposed for integrating the concept of time into Semantic Web applications, but they differ in various ways, e.g in the dimensions of

ISWC2023: International Semantic Web Conference, November 06–10, 2023, Athens, Greece
oleksandra.bruns@fiz-karlsruhe.de (O. Bruns)
0000-0002-8501-6700 (O. Bruns)

time they adapt, the type of temporal data considered, and how temporal statements in data can be expressed, which influences how the information is accessed and queried. For instance, some approaches consider the valid time to represent the validity of a certain fact, concept, or event, while others record the time of change in a KG as transaction time. Thus, the choice of a certain modeling approach is highly dependent on the use case. Additionally, historical data often lack complete provenance, making it challenging to document and represent the validity of a fact or concept accurately. To summarize, a structured and formal representation of temporal context through semantically interlinked representations such as KGs can provide novel means of exploring historical knowledge. However, the modeling approach must be carefully chosen based on the context and nature of the historical data to ensure that the temporal statements in the data are accurately represented and reasoned.
This research addresses the challenge by answering the research questions:

RQ1. What are the current best practices for representing temporal information in RDF? What are the key extensions to RDF that are currently being used to represent temporal information? What are their advantages and limitations?

RQ2. What temporal information is crucial in the context of cultural heritage? What requirements can be deduced from cultural heritage use cases?

RQ3. How can RDF extensions be evaluated? What characteristics are crucial for modeling time for cultural heritage use cases?

RQ4. How can RDF be applied to reason over temporal relationships? For example, how can it be inferred that one event occurred before another based on their timestamps?

RQ5. Can clear domain requirements and lessons learnt being utilized to develop a new lightweight and intuitive approach to model time in RDF? What criteria have to be considered?

Answering these research questions will have a significant impact on the cultural heritage field and beyond. Benefits may include improved organization and understanding of historical data by cultural heritage domain experts, including events, timelines, and narratives. This can lead to new insights and discoveries about the past. Additionally, a structured and clear view on how time can be represented in RDF can boost the development of new digital tools, e.g. for analyzing historical events or for visualizing interactive timelines of historical events; means for representation and exploration of historical processes, e.g. court processes; and the overall interoperability within heterogeneous tangible and intangible cultural data, e.g. historical events and the archival objects that recorded them, dramas and their performances and interpretations over time, etc. Ultimately, a better understanding and improvement of temporal data representation will contribute to enhancing public engagement with cultural heritage, and increase cross-disciplinary collaboration among researchers.

The reminder of the paper is structured as follows: Section 2 answers RQ1 and presents the work related to representing time in RDF and CH. Section 3 reports the preliminary results: Section 3.1 addresses RQ2 and derives modeling requirements, while Section 3.2 describes the evaluation plan (RQ3-4). Section 4 concludes the work and gives the insight to the future of the research (RQ4-5).

2. Related Work

There are two main approaches of representing and organizing knowledge by means of ontologies: event-based and entity-based, with the entity (object) and event as the main building blocks respectively [1]. In entity-based modeling, standard RDF triples represent statements about entities at a particular moment, without any explicit reference to the temporal context in which those statements hold. Which makes it impossible to reason over temporal information, while the event-based approach is developed to capture the dynamic nature of the data [2]. For this, a so-called event reification approach is used, which introduces a new special event entity for every change of state that may be associated with, inter alia, temporal context. One example is the CIDOC-CRM - the ISO standard ontology for cultural heritage domain [3], and is commonly used as underlying data model for CH KGs for representing national heritage across cultural disciplines, e.g. ArCo[4], and Sampo use cases[5], as well as domain specific knowledge, e.g. SeaLiT[6] is a KG of maritime history from the 1850s to the 1920s, and Odeuropa[7] to model olfactory heritage information and the semantic of smell. Following the event-based approach, CIDOC-CRM associates temporal information to the entities of the type E5 Event[1], e.g. *the death of Friedrich Schiller, the construction of the Berlin Wall*, etc. This allows for associating event entities with time directly, however results in very complex structures that require a significant level of expertise and effort to understand, implement and query correctly.

This research focuses on temporal extensions to RDF that allow for modeling temporal information in an entity-based setting. For the sake of generalisation, with regard to the level where the temporality is introduced the existing work towards extending RDF with time can be classified into a) graph level extensions, b) triple level extensions, c) component level extensions.

Graph level extensions are used to group RDF triples that share temporal entailment. For example, in Named Graphs [8] and n-quads[9] the quad format $< s, p, o, NamedGraph >$ is adopted, where the *NamedGraph* is linked to the temporal validity of triples. *Triple level extensions*, e.g. reification, tRDF [10, 11, 12, 13], aRDF[14], RDF* [15], either explicitly convert a triple into a reified statement or implicitly attach a temporal label to it. In *component level extensions*, the time label is attached to one or more triple components. For example, 4D fluents [16] convert a subject or an object of a triple into their time slices that carry temporal data. In the case of Singleton property approach [17], every predicate is unique and exists only in a certain temporal context. N-ary relations[18] objectify the relation and annotate it with its validity in time.

Adding a temporal dimension to RDF requires the introduction of temporal semantics to represent concepts such as time instants, intervals, durations, and relationships between them, as well as to reason over them. Some extensions, e.g. tRDF and 4D fluents, extend RDF semantics and propose temporal vocabularies. Others, in particular extensions that were developed for more generic purposes, e.g. Named Graphs and RDF*, require the use of time ontologies to incorporate temporal semantics, e.g. OWL-Time[2], BioTop[3], the Timeline Ontology[4], PROV-O[5]. In [19], a comprehensive review and comparison of the existing extensions to RDF were con-

---

[1] https://www.cidoc-crm.org/Entity/e5-event/version-6.2
[2] https://www.w3.org/TR/owl-time/
[3] https://bioportal.bioontology.org/ontologies/BT/
[4] https://motools.sourceforge.net/timeline/timeline.html
[5] https://www.w3.org/TR/prov-o/

ducted. The approaches were evaluated against several characteristics, as e.g., the need for additional objects, the number of triples required for temporal annotation, the extension of triples with additional dimensions, the specification of a query language, etc. However, based on the evaluation no clear guidelines could be defined since the choice of the model is highly use case specific. For example, in [20], the authors emphasize the significance of modeling temporal information for historic research, but note that practical suitability of the existing approaches towards it is still uncertain. To the best of our knowledge, no research has been conducted towards the best practices of representing temporal information for CH domain. This presented research task aims at addressing the gap by, first, defining clear requirements for modeling time in CH, and, second, evaluating the approaches against these requirements to provide modeling guidelines.

3. Towards Time Modeling in Cultural Heritage

This section aims at reporting the conducted work and preliminary results, in particular by defining modeling requirements for temporal information in CH, proposing specific characteristics for evaluation of the existing approaches, and creating an evaluation plan.

3.1. Modeling Requirements

In [21, 22, 23, 24, 25], the first steps towards the definition of the modeling requirements for CH based on real use cases, developed within TRANSRAZ [6], Wiedergutmachung [7], and "Subject Related Points of Access within Archivportal-D on Example of the subject area Weimar Republic"[8] projects, were conducted. For this, the bottom-up approach was followed, based on the specific nature of the data and competency questions (CQs) defined by the domain experts, e.g. archivists or historians, generic modeling requirements (REQs) were deduced:

- REQ1. **Conceptual evolution of entities** refers to cases where the meaning of entities changes over time, such as occupations undergoing a shift in their definition or perception.
- REQ2. **Terminological changes of entities** occur when labels associated with entities undergo modifications, e.g. change of a last name due to marriage or renaming of a street.
- REQ3. **Temporal annotation of entities** is crucial in a cultural context and involves associating specific time information with entities, such as birth and death dates of individuals, the establishment and demolition dates of historical buildings, or the formation of art groups.
- REQ4. **Temporal annotation of facts** involves assigning time validity to facts based on the changing relationships between entities over time. For example, the address of a created work may change.
- REQ5. **Modeling time stamps and intervals** is necessary due to the diverse nature of cultural data, where the time validity of entities and facts may be indicated in various ways, e.g. "The Marriage of Figaro (1786)", "Mozart's Last Solo (1791-03-04)".

[6] https://www.fiz-karlsruhe.de/en/forschung/transraz
[7] https://www.fiz-karlsruhe.de/en/forschung/wiedergutmachung
[8] https://www.fiz-karlsruhe.de/en/forschung/archivportal-d-sachthematische-zugaenge

REQ6. **Modeling of anonymous intervals** indicates intervals with incomplete or unknown temporal boundaries, e.g. "until 1923".
REQ7. **Modeling time uncertainty** is crucial for annotations where the time validity is contradictory or unclear, such as cases where approximate dates are provided, like "1908 or 1909," or when a rough estimation is given, such as "around 1200".
REQ8. **Modeling Allen's interval relations** allows for the exploration of hidden timelines and the extension of existing knowledge. This involves representing temporal relationships using Allen's interval algebra, enabling the identification of temporal order, overlap, and other relations, such as "during his studies" or "before that."

For example, from a specific CQ "Where were schools located in Nuremberg in 1908? How did their establishment develop throughout time?", several modeling requirements could be extracted: REQ3, since schools and addresses may change over time and have to be associated with their existence in time. REQ4, since schools may move and change their addresses, REQ2, e.g. streets or schools may be renamed, REQ1 e.g. a school with the same name can change it is function, REQ5, e.g. 1908, REQ8, e.g. the school should have been built before 1908.

3.2. Evaluation Plan and Criteria

After defining the requirements, the next research step involves evaluating existing approaches for time modeling based on these requirements. For this, we extend the list of criteria (CR) in [19] for the specific needs of the CH domain:

CR1. **Number of triples used for modeling a requirement**: This criterion aims to assess the efficiency and scalability of the models. Minimizing the number of triples required for modeling is important to avoid scalability issues when dealing with large datasets.
CR2. **RDF syntax extension**: Many models extend standard RDF triples to incorporate the temporal dimension. This characteristics considers the influence of such extensions on reasoning, the choice of triple store, etc.
CR3. **Extension of RDF(S)/OWL semantics**: It focuses on the extent to which the models extend the semantics of RDF(S) or OWL to add a temporal dimension. Evaluating the semantic extensions helps understand how well the models capture temporal information and reasoning capabilities.
CR4. **Use of native SPARQL for querying**: The ability to query the data using native SPARQL is an essential criterion for evaluating the usability and accessibility of the models.
CR5. **Expressivity of model-defined temporal semantics**: This criterion examines the richness and expressiveness of the temporal semantics against each requirement. Evaluating expressivity helps to determine the level of detail and complexity that can be captured when modeling temporal information. In particular, if the model is capable to represent a requirements, e.g. intervals, time stamps, interval relations, etc.
CR6. **Use of OWL reasoning**: This criterion explores the use of OWL reasoning to derive new data and observe the influence of an RDF extension on reasoning capabilities. Assessing the impact of OWL reasoning on the modeled temporal data provides insights into the models' inferencing capabilities.

CR7. **User-friendliness of a model**: Considering the cultural heritage domain, it is crucial to bear in mind that the potential users of the systems are often domain experts with little to no experience in RDF and querying. Therefore, an additional important criteria refers to nativity of each modeling approach. This criterion focuses on evaluating how intuitive and user-friendly the models are for domain experts, ensuring that they can effectively utilize the models without extensive querying knowledge.

In the next steps of the research, the existing models are evaluated based on the pre-defined characteristics. For this, sample RDF data from the use cases for each REQ are created using the existing RDF extensions described in Section 2. At the moment of writing the evaluation is ongoing. Several criteria are being considered, including both general aspects of the model and specific requirements. Regarding the model in general, CR2 examines whether the model utilizes the abstract syntax of RDF and identifies the concrete syntaxes employed, such as TTL, TriG, or TTL*, since this effects the model usability. CR3 reports the underlying semantics on which the extension is built on, e.g. RDF(S), OWL, annotated logic, etc. CR4 determines if the SPARQL is sufficient for querying the graph or what alternative query languages are utilized.
For each model, the remaining criteria are evaluated against the defined requirements to identify strengths and weaknesses. For CR1, the number of triples/quads to model a requirement is counted. CR5 evaluates if the extension enables expressive representation of a requirement, e.g. if it allows for modeling time intervals or interval relations. For models that do not have a specified time ontology, it is important to incorporate external vocabularies like RDF*+Owl-Time or RDF*+TimeLine to establish a suitable framework. To evaluate the user-friendliness of the model, CR7, user testing sessions are required, where the feedback of representative users on querying usability is to be collected.
For CR6, the evaluation criteria to compare the OWL inference must be considered, since a simple binary answer is insufficient. OWL inference operates on ontologies represented by triples, while, e.g. Named Graphs do not have a standartized quad semantics and simply provide a way to organize and label triples [8]. Although some triple stores support OWL inference over Named Graphs, the reasoning process does not directly consider the origins of triples within Named Graphs. Similarly, tRDF (temporal reification) allows for OWL inference, but the effectiveness of OWL language constructs becomes significantly restricted when dealing with reified relationships. This limitation arises because much of OWL's inference power is dedicated to describing binary relations, such as cardinality, inverses, transitivity, etc. However, when treated as classes, the OWL reasoning over relations becomes limited.

4. Conclusion and Future Work

The presented research provides a significant contribution to the representation of temporal information in cultural heritage knowledge graphs, addressing the critical yet challenging aspect of incorporating temporality specifically for this domain. By defining specific modeling requirements based on real use cases and proposing evaluation characteristics, it enables a more user-oriented analysis and comparison of existing work in modeling temporal knowledge. It is argued that evaluation of the approaches against individual requirements allows for a more clear identification of models' strengths and drawbacks. The process of requirement definition

is iterative and adaptable to address and fit to new use cases as they arise.
In the next stages of the research, the focus will shift towards addressing RQ4 and RQ5. This involves conducting evaluations of the existing work and presenting a comprehensive analysis. The results obtained will then be utilized to derive guidelines for CH practitioners. Furthermore, based on the insights gained from the evaluations and analysis, goal is to develop a new intuitive and lightweight approach for modeling and reasoning over temporal data. This approach will leverage the lessons learned to create an effective solution that is user-friendly and well-suited for the unique challenges and complexities of the CH domain.
Ultimately, the findings of this research will pave the way for advancements in time modeling approaches within the cultural heritage field. By addressing the specific needs and requirements of CH practitioners, it will contribute to the overall improvement and progress in handling temporal knowledge in cultural heritage contexts.

Acknowledgments

This work is supervised by Prof. Dr. Harald Sack.

References

[1] G. N. Allen, S. T. March, The effects of state-based and event-based data representation on user performance in query formulation tasks, Mis Quarterly (2006) 269–290.
[2] J. P. A. Almeida, R. A. Falbo, G. Guizzardi, Events as entities in ontology-driven conceptual modeling, in: Conceptual Modeling: 38th Int. Conf., ER 2019, Salvador, Brazil, November 4–7, 2019, Proc. 38, Springer, 2019, pp. 469–483.
[3] M. Doerr, The CIDOC CRM, an ontological approach to schema heterogeneity, in: Dagstuhl Seminar Proc., Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2005.
[4] V. A. Carriero, A. Gangemi, M. L. Mancinelli, L. Marinucci, A. G. Nuzzolese, V. Presutti, C. Veninata, Arco: The italian cultural heritage knowledge graph, in: The Semantic Web–ISWC 2019: 18th Int. Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18, Springer, 2019, pp. 36–52.
[5] E. Hyvönen, "sampo" model and semantic portals for digital humanities on the semantic web, in: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020), CEUR-WS. org, 2020.
[6] A. Delis, Seafaring lives at the crossroads of mediterranean maritime history, Int. Journal of Maritime History 32 (2020) 464–478.
[7] P. Lisena, D. Schwabe, M. van Erp, R. Troncy, W. Tullett, I. Leemans, L. Marx, S. C. Ehrich, Capturing the semantics of smell: The odeuropa data model for olfactory heritage information, in: European Semantic Web Conference, Springer, 2022, pp. 387–405.
[8] J. J. Carroll, C. Bizer, P. Hayes, P. Stickler, Named Graphs, Provenance and Trust, in: Proc. of the 14th int. Conf. on World Wide Web, 2005, pp. 613–622.
[9] G. Carothers, RDF 1.1 N-Quads: A line-based syntax for RDF datasets, W3C recommendation 25 (2014).

[10] C. Gutierrez, C. Hurtado, A. Vaisman, Temporal rdf, in: European Semantic Web Conf., Springer, 2005, pp. 93–107.

[11] C. Gutierrez, C. A. Hurtado, A. Vaisman, Introducing time into RDF, IEEE Transactions on Knowledge and Data Engineering 19 (2006) 207–218.

[12] C. Hurtado, A. Vaisman, Reasoning with temporal constraints in RDF, in: Int. Workshop on Principles and Practice of Semantic Web Reasoning, Springer, 2006, pp. 164–178.

[13] A. Pugliese, O. Udrea, V. Subrahmanian, Scaling RDF with time, in: Proc. of the 17th Int. Conf. on World Wide Web, 2008, pp. 605–614.

[14] O. Udrea, D. R. Recupero, V. Subrahmanian, Annotated rdf, ACM Transactions on Computational Logic (TOCL) 11 (2010) 1–41.

[15] O. Hartig, Foundations of RDF* and SPARQL*:(An alternative approach to statement-level metadata in RDF), in: AMW 2017 11th Alberto Mendelzon Int. Workshop on Foundations of Data Management and the Web, Montevideo, Uruguay, June 7-9, 2017., volume 1912, Juan Reutter, Divesh Srivastava, 2017.

[16] C. Welty, R. Fikes, S. Makarios, A reusable ontology for fluents in OWL, in: FOIS, volume 150, 2006, pp. 226–236.

[17] V. Nguyen, O. Bodenreider, A. Sheth, Don't like RDF reification? Making statements about statements using singleton property, in: Proc. of the 23rd Int. Conf. on World wide web, 2014, pp. 759–770.

[18] N. Noy, A. Rector, P. Hayes, C. Welty, Defining n-ary relations on the semantic web, W3C working group note 12 (2006).

[19] H.-T. Wang, A. U. Tansel, Temporal extensions to RDF, Journal of Web Engineering 18 (2019) 125–168.

[20] A. Meroño-Peñuela, A. Ashkpour, M. Van Erp, K. Mandemakers, L. Breure, A. Scharnhorst, S. Schlobach, F. Van Harmelen, Semantic technologies for historical research: A survey, Semantic Web 6 (2015) 539–564.

[21] O. Bruns, T. Tietz, M. Vafaie, D. Dessì, H. Sack, et al., Towards a representation of temporal data in archival records: Use cases and Requirements, in: CEUR Workshop Proc., volume 3019, CEUR-WS, 2021, pp. 128–134.

[22] O. Bruns, T. Tietz, M. B. Chaabane, M. Portz, F. Xiong, H. Sack, The Nuremberg Address Knowledge Graph, in: The Semantic Web: ESWC 2021 Satellite Events: Virtual Event, June 6–10, 2021, Revised Selected Papers 18, Springer, 2021, pp. 115–119.

[23] O. Vsesviatska, T. Tietz, F. Hoppe, M. Sprau, N. Meyer, D. Dessì, H. Sack, ArDO: an ontology to describe the dynamics of multimedia archival records, in: Proc. of the 36th Annual ACM Symposium on Applied Computing, 2021, pp. 1855–1863.

[24] T. Tietz, O. Bruns, S. Göller, M. Razum, D. Dessì, H. Sack, Knowledge Graph enabled Curation and Exploration of Nuremberg's City Heritage, in: Qurator 2021 Conf. on Digital Curation Technologies, Berlin, Germany, February 8th to 12th, 2021, volume 2836, Aachen, Germany: RWTH Aachen, 2021, p. 11.

[25] M. Vafaie, O. Bruns, N. Pilz, D. Dessí, H. Sack, Modelling Archival Hierarchies in Practice: Key Aspects and Lessons Learned, in: 6th Int. Workshop on Computational History (HistoInformatics 2021), co-located with ACM/IEEE Joint Conf. on Digital Libraries 2021 (JCDL 2021), Online event, September 30-October 1, 2021, volume 2981, Aachen, Germany: RWTH Aachen, 2021, p. 6.