

CourtDocs Ontology: Towards a Data Model for Representation of Historical Court Proceedings

Mahsa Vafaie
FIZ Karlsruhe – Leibniz Institute for
Information Infrastructure
Karlsruhe Institute of Technology
Karlsruhe, Germany
mahsa.vafaie@fiz-karlsruhe.de

Oleksandra Bruns
FIZ Karlsruhe – Leibniz Institute for
Information Infrastructure
Karlsruhe Institute of Technology
Karlsruhe, Germany
oleksandra.bruns@fiz-karlsruhe.de

Nastasja Pilz
State Archives of Baden-Württemberg
Stuttgart, Germany
nastasja.pilz@la-bw.de

Jörg Waitelonis
FIZ Karlsruhe – Leibniz Institute for
Information Infrastructure
Karlsruhe Institute of Technology
Karlsruhe, Germany
joerg.waitelonis@fiz-karlsruhe.de

Harald Sack
FIZ Karlsruhe – Leibniz Institute for
Information Infrastructure
Karlsruhe Institute of Technology
Karlsruhe, Germany
harald.sack@fiz-karlsruhe.de

ABSTRACT

For several decades researchers have studied legal documents for insights into the evolution of legal norms and strategies, in their social and cultural context. Analysing these documents and the associated legislative sessions, trials and court cases helps uncover hidden narratives and patterns, as well as showcase the lessons learnt. The field of knowledge engineering has contributed to the growing interest in the development and use of legal ontologies that aim at providing machine-readable foundations to model legal concepts, relations and processes. Legal ontologies have been used for legal knowledge management and as knowledge bases in legal knowledge systems. With a focus on the *Wiedergutmachung* project as a use case, this paper presents an overview of the existing legal ontologies, demonstrates the gap to align them with the essential conceptual framework required to model historical court proceedings with respect to provenance information, and presents the ongoing work towards developing the CourtDocs Ontology by utilising existing standards and ontologies on the intersection of the legal domain, history and archival sciences. The *Wiedergutmachung* project centres around constructing a knowledge graph as a backbone for information systems, based on historical archival records from the compensation procedure in post-World War II Germany.

CCS CONCEPTS

• **Information systems** → **Digital libraries and archives: Ontologies.**

KEYWORDS

Archival Documents, Knowledge Representation, Ontology

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

K-CAP '23, December 05–07, 2023, Pensacola, FL, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0141-2/23/12...\$15.00

<https://doi.org/10.1145/3587259.3627562>

ACM Reference Format:

Mahsa Vafaie, Oleksandra Bruns, Nastasja Pilz, Jörg Waitelonis, and Harald Sack. 2023. CourtDocs Ontology: Towards a Data Model for Representation of Historical Court Proceedings. In *Knowledge Capture Conference 2023 (K-CAP '23), December 05–07, 2023, Pensacola, FL, USA*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3587259.3627562>

1 INTRODUCTION

Law, legal documents, and legal procedures are interconnected components of a comprehensive legal system. They are essential for maintaining order, protecting rights, resolving conflicts, and promoting justice. Law and legal procedures have existed for thousands of years, evolving and adjusting in response to a dynamic world, while reflecting cultural and social dimensions of human existence. Legal documents serve as tools to record the abstractions of law, documenting procedures, actions, timelines, and criteria that must be followed to comply with the rules of law. Analysing these documents alongside the corresponding legal procedures aids in revealing hidden narratives and recurring patterns, while also highlighting valuable insights gained from these experiences. One means to organise, represent and analyse legal information is through the utilisation of legal ontologies. Legal ontologies are structured representations of legal knowledge, encompassing concepts, definitions, and relationships within the legal domain. They are frequently employed by researchers, practitioners, and policy-makers to effectively navigate the complexities of legal documents and procedures. For instance, UK Legislation¹ aggregates legislation from across the United Kingdom using the MetaLex ontology [3]. A set of ontologies has been modelled to represent different aspects of semantics of the legal domain. For instance, FOLaw [6] aims at clarifying different types of knowledge and their interconnections in the domain, e.g., causal knowledge and responsibility knowledge. Efforts to represent and conceptualise entities obtained from legal texts, e.g., agents, their roles, and events are presented by LRI-Core [5] and LKiF [15]. Further studies focus on modelling various legal elements, such as judicial acts (CLO [12]), legal relations (UFO-L [13]), legal rules (LegalRuleML [2]), and legislative

¹<https://www.legislation.gov.uk/>

projects (ELI Ontology [11]). Subsequent ontologies, each designed to formalise distinct facets of law, have been thoroughly examined in the survey presented in [10].

Yet, the knowledge embedded in archival legal documents is also of great interest to archivists, historians and the general public. Archival legal documents, vital for historical research, pose challenges to existing legal ontologies. Firstly, these documents are cultural heritage objects, thus, are often incomplete or inaccurate. Secondly, historical legal documents may refer to historical events, contexts, and entities that require specialised historical and cultural knowledge to interpret and represent. Finally, each legal procedure carries distinct attributes and characteristics. However, conventional legal ontologies are designed top-down to fit broader common sense and legal principles, which may result in lacking the level of granularity required for precise representation of a specific application.

Structuring and contextualising historical knowledge for historical exploration and analysis, and creation of an ontology-based information system are the objectives of the project “Themenportal Wiedergutmachung”^{2,3}. This project is centred around archival documents from the process of compensation for injustices that occurred during the rule of the Nazi regime. Wiedergutmachung archival records originate from the State Offices for Compensation (*Ämter für Wiedergutmachung* in German) installed by the German government in every federal state after World War II. Based on the granted archival documents, the endeavour aims to reconstruct the compensation processes and provide comprehensive resources for understanding the societal transformation after 1945.

This paper presents the ongoing work towards developing an application ontology for the use case of Wiedergutmachung, while adhering to FAIR principles and practices in ontology development [20] that allow for construction of a generic ontology that could be reused in similar contexts. CourtDocs Ontology aims at representing historical court proceedings; it utilises existing vocabularies and focuses around three central concepts:

- **Archival documents** record reparation processes and cases, e.g., application form, notice of compensation.
- **Legal Processes** are activities that are performed to enforce a law, e.g., application, investigation, decision.
- **Agents** are parties that are involved in a process, e.g., applicants, victims, lawyers, State Offices for Compensation.

In the following, the ongoing work on CourtDocs ontology, in particular the definition of competency questions, modelling requirements, reused ontologies and a discussion of the ontological design for the Wiedergutmachung use case is presented.

2 MODELLING REQUIREMENTS FOR COURTDOS ONTOLOGY

Developing an information system on the basis of historical archival documents hinges upon two primary tasks: creation of an ontology for knowledge representation and extraction of information from the documents to build up a knowledge graph, as a foundation for semantic and exploratory search. An ontology serves as a collective

conceptual framework and the rational foundation for the knowledge graph, outlining the associations among ideas and concepts [14]. Engaged users and domain experts determine the domain and scope of an ontology, with the consideration of the inquiries they aim to resolve. In this use case, the archivists at State Archives of Baden-Württemberg⁴, which holds a portion of the archival data, collaborated with the authors to formulate a comprehensive set of competency questions, as the foundational basis for structuring the CourtDocs Ontology. They shared their proficiency on the Wiedergutmachung process and their familiarity with potential users of the Wiedergutmachung Knowledge Graph. When crafting the inventory of competency questions, two distinct target audiences were considered: The first group includes researchers and historians with an interest in the subject matter, while the second group comprises the relatives and dependents of those who have applied for compensation. The complete list of competency questions can be found on GitHub in the Wiedergutmachung repository⁵. The competency questions can potentially be categorised into four distinct groups, depending on the subject they inquire about:

- Document Search: e.g., *Do we find proof of citizenship inside the documents?*
- Biographical Research: e.g., *Do we find the famous lawyer Dr. Benno Ostertag (13.09.1892) in the files? Who did he plead cases for?; How many victims survived the camps in Auschwitz/Dachau/etc.?*
- Prosopographical Research: e.g., *Can we identify cases for certain professions?; How did the Roma people or Jehova's Witnesses suffer from Nazi persecution?*
- Court case specifications:
 - Damage types: e.g., *What are significant categories of damages to health and body?*
 - Compensation sums: e.g., *Were the reparation sums higher when the victim had a lawyer?*
 - Role-based search: e.g., *What lawyers frequently appear in the process? How did they influence the process?*

As previously indicated, CourtDocs Ontology is designed to comprise three primary foundational cores: archival documents, legal processes, and agents. For modelling each of these components, existing ontologies are reused to prevent redundancy and, in turn, to foster compatibility with external data sources. This modular design and the modelling requirements for each of these foundational components are drawn from the competency questions, discussions and interviews with the archivists, and guided by established principles in the domain of Ontology Design [1]. In the subsequent sections we expound on the modelling requirements and the ontologies that have been reused for representation of each component of CourtDocs Ontology.

2.1 Representation of Archival Documents

At the centre of “Themenportal Wiedergutmachung” are collections of documents, records, and materials related to the process of compensation that followed the fall of the Nazi regime. The Wiedergutmachung compensation records are held in archival institutes all

²<https://www.archivportal-d.de/themenportale/wiedergutmachung>

³https://is.gd/bundesfinanzministerium_wgm

⁴<https://www.landesarchiv-bw.de/>

⁵<https://github.com/ISE-FIZKarlsruhe/Wiedergutmachung/>

over Germany. Capturing interrelatedness of archival data and detailed provenance information as well as facilitating description, search and navigation of archival records is enabled by modelling the archival documents according to formal representations [18]. The requirements for modelling historical archival documents in the context of court procedures are drafted below.

- REQ 1.1:** Historical archival documents are considered cultural heritage objects and require semantic representation for findability and interlinking with external data sources.
- REQ 1.2:** Provenance information is required to keep track of the origin of statements and entities, and sources they were derived from.
- REQ 1.3:** Archival documents are represented hierarchically in the archiving system. Modelling this hierarchical structure is necessary for an accurate representation of the documents.
- REQ 1.4:** Archival documents have been used by and generated from different stages of the court procedure. In order to facilitate document search and document-specific information retrieval it is necessary to extend the archival ontology in such a way that the relations between archival documents and the court procedure are captured.

To fulfill all these modelling requirements, the Records in Contexts-Ontology (referred to as RiC-O⁶) [9] has been adopted. [18] present an overview of other archival ontologies and a detailed discussion on the rationale behind selecting RiC-O for this specific use case. To summarise, this choice is mainly motivated by RiC-O's flexible structure and level of granularity. Despite the implementation of archival standards and principles such as ISAD(G)[8] and the so-called "Basic Principle of Arrangement"⁷, categorisation of archival documents into smaller subgroups (such as fonds, series, collections, etc.) and the extent of detail in their description is always contingent on the nature of the material, the archival institute and the archivist managing them. For example, some archives allow fonds to be parts of series, whereas in other archives this is not allowed. In RiC-O, the hierarchy is expressed by bounding archival resources to a set of named individuals, to describe which categorisation scheme is used to classify the archival resources. These named individuals are connected to *rico:RecordSet* with the property *rico:hasRecordSetType* and loosely linked to one another through the property *rico:includesOrIncluded*. This feature enhances the flexibility of the model, enabling its application across various archival collections from different archival institutes and along different periods of time, even when dealing with non-standardised archival systems and practices.

Additionally, RiC-O's effectiveness is augmented by its ability to include entities smaller than an archival record within the hierarchy. A core entity in RiC-O is *rico:RecordResource*, with three sub-classes, *rico:RecordSet*, *rico:Record* and *rico:RecordPart*. According to the documentation, "determining when an information object is a Record, Record Part, or Record Set is based on perspective and judgement exercised in a particular context". The inclusion of the class *rico:RecordPart* enables a more granular representation of

archival materials and their components, such as archival annotations and stamps on documents [18].

Finally, in order to address REQ 1.4, RiC-O is extended with newly defined subclasses for the class *rico:Record*, such as *cd:ApplicationForm* and *cd:SupplementaryRecord*. This increased level of granularity acknowledges the legal characteristics of the archival records and is intended for reuse in other legal use cases.

2.2 Modelling the Legal Processes

While archival records serve as the primary source of information and are central to CourtDocs Ontology, developing a model that represents every individual document is inefficient and labour-intensive. The composition of documents within each Wiedergutmachung case file varies according to the specific case, and the data extractable from these documents can expand based on the users' primary areas of interest. Nevertheless, specific documents consistently appear in all case files, playing a crucial role as they either initiate or result from activities essential to the Wiedergutmachung process; namely *Application*, *Investigation*, *Decision*, and *Compensation Payment*, as illustrated in the figure for CourtDocs data model published in the Wiedergutmachung GitHub repository⁸. Termed as "key documents" within this work, these documents are associated with specific events and serve the purpose of reconstructing the legal proceedings of Wiedergutmachung. They represent the only available account of how the compensation process unfolded for each individual application case. The requirements for modelling the legal process of Wiedergutmachung, which are equally relevant for other compensation processes, are listed below.

- REQ 2.1:** Each court procedure consists of consecutive events. These events can be categorised into event types that might reoccur multiple times (e.g., investigation).
- REQ 2.2:** The reconstruction of the timeline for these events is possible with the help of archival documents. The archival documents represent the events and activities that constitute the court procedures.
- REQ 2.3:** Agents with certain attributes, e.g., occupations and group memberships, participate in each of the events. Linking these agents with events that constitute the legal processes and corresponding documents should enable exploration of the role these agents play in the court procedure.

In the domain of legal ontologies, LRI-Core [5] provides the possibility to model agents and events; however, it is an upper ontology which does not fit the level of granularity needed for this use case, and to the knowledge of the authors, is not currently maintained. On the other hand, the event-based data model CIDOC-CRM is designed to capture a wide range of detailed information, including events and provenance. However, its level of detail and model complexity often results in inefficiencies when querying and processing data [7]. Working with this data model requires a thorough understanding of its structure and principles, which may pose challenges to the potential future users of the ontology, seeking effective adoption and utilisation.

⁶<https://www.ica.org/standards/RiC/ontology>

⁷<https://www.archives.gov/research/alic/reference/archives-resources/principles-of-arrangement.html>

⁸<https://github.com/ISE-FIZKarlsruhe/Wiedergutmachung/tree/main/CourtDocsOntology>

The PROV Ontology (PROV-O⁹) [16] is used for modelling the process of Wiedergutmachung within the legal realm. PROV-O effectively functions as an ontology for modelling processes, providing a comprehensive and standardised methodology for delineating the evolution of and interrelations between entities, activities and agents over time within a system. Conversely, PROV-O is specifically designed to portray the provenance linked to the creation of distinct data elements or entities, taking into account both the activities and agents responsible. This inherent characteristic of PROV-O makes it exceptionally suitable for illustrating the associations between court procedures and the corresponding records generated or utilised at each phase of the procedure.

2.3 Modelling Agents

Agents play a crucial role in legal processes, facilitating various tasks and interactions on behalf of individuals and organisations. Understanding their roles and their involvement in historical legal processes helps to comprehend the dynamics of legal systems throughout history, to reconstruct the timelines and narratives of lives shaped by those legal landscapes, and to showcase how individuals and institutions act in different historical contexts. Within the Wiedergutmachung use case, the following requirements are identified for modelling agents:

REQ 3.1: Agents are interconnected via different relations. These relations have to be semantically represented.

REQ 3.2: Agents who participate in events can take up different roles in the events. Representation of their participation role is important for role-based search.

REQ 3.3: Descriptions of agents are often extracted from historical resources, thus, may be uncertain, incomplete or even false. Therefore, provenance information is required to confirm authenticity.

REQ 3.4: Attributes of agents, e.g., their addresses, names, occupations, etc. may change over time.

Numerous ontologies have been developed to address various aspects of agent description from distinct perspectives. General ontologies, e.g., FOAF¹⁰, DBO¹¹, Schema.org¹², and BFO¹³ have been developed with different objectives and contain a rich semantic representation and hierarchy of social relations between individuals. However, they lack concepts for modelling historical and legal aspects of the use case. Additionally, the dynamic attributes of agents and their temporal association with legal processes (e.g., that roles may change within the process) require an expressive modelling of temporal aspects, which is beyond the scope of the aforementioned ontologies.

To represent biographical information about individuals engaged in the compensation process comprehensively, a combination of ontologies proves most fitting. For modelling historical agents in CourtDocs Ontology, PROV-O is used to handle provenance and participation roles, while OWL-Time¹⁴ tackles temporal considerations.

Despite CIDOC-CRM's¹⁵ [19] reputation for introducing complexity and query inefficiencies due to its event-based nature [7], its strength lies in capturing historical and cultural contexts, making it a popular choice within GLAM institutions (Galleries, Libraries, Archives, Museums). To enhance interoperability, CourtDocs links to, and reuses concepts from CIDOC-CRM to represent the historical and cultural contexts of agents involved in the Wiedergutmachung process. Furthermore, an efficient approach for harmonising and structuring biographical information involves an event-centric modelling such as CIDOC-CRM, wherein an individual's existence can be envisioned as series of interconnected incidents spanning time and space, commencing with birth and concluding with death [17]. This approach facilitates the representation of supplementary life occurrences that bear significance in prosopographical research concerning victims of the National Socialism era, such as instances of deportation and imprisonment.

3 CONCLUSIONS AND FUTURE WORK

This paper highlights ongoing efforts to develop an application ontology tailored to the Wiedergutmachung use case, adhering to FAIR principles [20] and user-centered ontology design practices [4]. The unique challenges that historical archival records documenting court proceedings present to conventional legal ontologies are explained in the context of the Wiedergutmachung use case. While existing legal ontologies are primarily designed to represent broad legal principles and norms, they lack the granularity required for precise representation of legal processes in specific applications. The CourtDocs Ontology, which focuses on representing historical court proceedings, utilises existing vocabularies while centralising around three fundamental concepts: archival documents, legal processes, and agents involved in these processes. While designed for the specific use case of Wiedergutmachung, one of the main considerations in development of CourtDocs has been its reusability in similar contexts but with different specifications.

A comprehensive evaluation of this ontology for the current and potential future use cases is yet to be completed, in order to assess the modelling power of CourtDocs Ontology. Moreover, it is crucial for this ontology to stay adaptable and extendable, due to its user-centric design, which necessitates future iterations for the expansion of the set of competency questions. Additionally, to enhance interoperability and foster interconnections with other data sources, it is pivotal to conduct a mapping of CourtDocs concepts with existing in-use ontologies and frameworks such as the European Legislation Identifier (ELI) Ontology. Further research and development of ontologies at the intersection of history and law can help facilitate a more nuanced understanding of historical legal processes and foster interoperability in the domains of historical and legal information systems.

ACKNOWLEDGMENTS

This work is funded by the German Federal Ministry of Finance (*Bundesministerium der Finanzen*).

⁹<https://www.w3.org/TR/prov-o/>

¹⁰<http://xmlns.com/foaf/0.1/>

¹¹<https://www.dbpedia.org/ontology/>

¹²<https://www.schema.org/>

¹³<https://basic-formal-ontology.org/>

¹⁴<https://www.w3.org/TR/owl-time/>

¹⁵<https://cidoc-crm.org/>

REFERENCES

- [1] Dean Allemang, James A Hendler, and Fabien Gandon. 2020. *Good and Bad Modeling Practices*. Association for Computing Machinery, New York, NY, USA, Chapter 15, 425–452. <https://doi.org/10.1145/3382097.3382113>
- [2] Tara Athan, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. 2015. *LegalRuleML: Design Principles and Foundations*. Springer International Publishing, Cham, 151–188. https://doi.org/10.1007/978-3-319-21768-0_6
- [3] Alexander Boer, Radboud Winkels, and Fabio Vitali. 2008. Metalex XML and the legal knowledge interchange format. In *Computable Models of the Law: Languages, Dialogues, Games, Ontologies*. Springer, Berlin, Heidelberg, 21–41.
- [4] Maricela Bravo, Luis Fernando Hoyos Reyes, and José A Reyes Ortiz. 2019. Methodology for ontology design and construction. *Contaduría y administración* 64, 4 (2019), 134.
- [5] Joost Breuker, Rinke Hoekstra, et al. 2004. Core concepts of law: taking common-sense seriously. In *Proceedings of formal ontologies in information systems (FOIS-2004)*. IOS Press, 210–221. <https://books.google.de/books?id=6-PLB7DiixAC>
- [6] Joost Breuker, Andre Valente, and Radboud Winkels. 2004. Legal Ontologies in Knowledge Engineering and Information Management. *Artificial Intelligence and Law* 12 (12 2004), 241–277. <https://doi.org/10.1007/s10506-006-0002-1>
- [7] George Bruseker, Nicola Carboni, and Anaïs Guillem. 2017. *Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM*. Springer International Publishing, Cham, 93–131. https://doi.org/10.1007/978-3-319-65370-9_6
- [8] Rainer Brüning, Werner Heegewaldt, and Nils Brübach. 2011. *ISAD(G) - General International Standard Archival Description, 2nd Edition (DE)*. Technical Report. International Council for Archives (ICA).
- [9] Florence Clavaud and Tobias Wildi. 2021. ICA Records in Contexts-Ontology (RiC-O): a Semantic Framework for Describing Archival Resources. In *Linked Archives 2021: Proceedings of Linked Archives International Workshop 2021 co-located with 25th International Conference on Theory and Practice of Digital Libraries (TPDL 2021)*, p. 79–92. <https://enc.hal.science/hal-03965776>
- [10] Cleyton Mario de Oliveira Rodrigues, Frederico Luiz Gonçalves de Freitas, Emanoel Francisco Spósito Barreiros, Ryan Ribeiro de Azevedo, and Adauto Trigueiro de Almeida Filho. 2019. Legal ontologies over time: A systematic mapping study. *Expert Systems with Applications* 130 (2019), 12–30.
- [11] Thomas Francart, John Dann, Roberto Pappalardo, Carmen Malagon, and Marco Pellegrino. 2019. The European legislation identifier. *Knowledge of the Law in the Big Data Age* 317 (2019), 137–148.
- [12] Aldo Gangemi, Alessandra Prisco, Maria-Teresa Sagri, Geri Steve, and Daniela Tiscornia. 2003. Some ontological tools to support legal regulatory compliance, with a case study. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. Springer, 607–620.
- [13] Cristine Griffo. 2015. Ufo-l: A core ontology of legal concepts built from a legal relations perspective. *Doctoral Consortium Contributions, IC3K-KEOD* (2015).
- [14] Thomas R Gruber. 1993. A translation approach to portable ontology specifications. *Knowledge acquisition* 5, 2 (1993), 199–220.
- [15] Rinke Hoekstra, Joost Breuker, Marcello Di Bello, Alexander Boer, et al. 2007. The LKIF Core Ontology of Basic Legal Concepts. *LOAIT* 321 (2007), 43–63.
- [16] Luc Moreau and Paul Groth. 2013. *The prov Ontology*. Springer International Publishing, Cham, 21–38. https://doi.org/10.1007/978-3-031-79450-6_3
- [17] Jouni Antero Tuominen, Eero Antero Hyvönen, and Petri Leskinen. 2018. Bio CRM: A data model for representing biographical data for prosopographical research. In *Proc. of the 2nd Conf. on Biographical Data in a Digital World 2017 (BD2017)*. CEUR Workshop Proceedings, CEUR-WS.org. <https://ceur-ws.org/Vol-2119>
- [18] Mahsa Vafaie, Oleksandra Bruns, Nastasja Pilz, Danilo Dessi, and Harald Sack. 2021. Modelling Archival Hierarchies in Practice: Key Aspects and Lessons Learned. In *6th International Workshop on Computational History (HistInformatics 2021), co-located with ACM/IEEE Joint Conference on Digital Libraries 2021 (JCDL 2021), Online event, September 30-October 1, 2021*, Vol. 2981. Aachen, Germany: RWTH Aachen, 6.
- [19] Athanasios Velios et al. 2021. Definition of the CIDOC Conceptual Reference Model.
- [20] Mark D. Wilkinson, Michel Dumontier, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.