

# A Data Model for Linked Stage Graph and the Historical Performing Arts Domain

Tabea Tietz<sup>1</sup>, Oleksandra Bruns<sup>1</sup> and Harald Sack<sup>1</sup>

<sup>1</sup>FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Germany

## Abstract

The performing arts are complex, dynamic and embedded into societal and political systems. Providing means to research historical performing arts data is therefore crucial for understanding our history and culture. However, currently no commonly accepted ontology for historical performing arts data exists. On the example of the Linked Stage Graph, this position paper presents the ongoing process of creating an application-driven and efficient data model by leveraging and building upon existing standards and ontologies like CIDOC-CRM, FRBR, and FRBRoo.

## Keywords

Cultural Heritage, Performing Arts, Knowledge Graphs, Research Data

## 1. Introduction

Theatre as an art form and social institution has been deeply embedded into our societies for thousands of years. Theatre is a realm that effortlessly merges the historic and the modern, continuously evolving and reinventing itself time and time again. Historical plays, which persist to this day, remain relevant, finding new life through reinterpretation on modern stages. Theatre never exists in isolation and is always embedded in the context of the respective culture and society. These contexts can be shaped by societal and political events or by technological advancements and these contexts also have shown to influence creative possibilities and limitations on stage. Scholars from historical sciences, theatre science, social science, and digital humanities have been examining the history and development of theater in all its facets, including stage design, characters, costumes, and language. Research questions being posed include:

- How did stage design develop over time for Shakespeare's *Midsummer Night's Dream*?
- Which actors were on stage during World War II in Germany?
- Which playwrights are most interpreted on modern stages?
- Which theatre companies produced the works of Bertolt Brecht during the late 1920s?


Several tools, databases, and platforms have been developed with the aim of granting access to collections of performing arts data. However, a significant portion of these initiatives employ


---

*SWODCH'23: International Workshop on Semantic Web and Ontology Design for Cultural Heritage, November 7, 2023, Athens, Greece*

✉ [tabea.tietz@fiz-karlsruhe.de](mailto:tabea.tietz@fiz-karlsruhe.de) (T. Tietz); [oleksandra.bruns@fiz-karlsruhe.de](mailto:oleksandra.bruns@fiz-karlsruhe.de) (O. Bruns); [harald.sack@fiz-karlsruhe.de](mailto:harald.sack@fiz-karlsruhe.de) (H. Sack)

ORCID  0000-0002-1648-1684 (T. Tietz); 0000-0002-8501-6700 (O. Bruns); 0000-0001-7069-9804 (H. Sack)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

closed systems or proprietary software, and they often refrain from publishing open APIs, data models, and data. This practice introduces barriers for the research community, impeding the discoverability, accessibility, interoperability, and reusability of performing arts research data.

One of these endeavors, the Linked Stage Graph, offers the opportunity to explore photographs and metadata from the Stuttgart State Theatres spanning the period from the 1880s to the 1940s [1]. The knowledge graph contains a range of performance events, encompassing theatre plays, ballets, and operas. The knowledge graph is accessible on the Web by means of a public SPARQL Endpoint<sup>1</sup> and an exploration interface<sup>2</sup>. However, the construction of this knowledge graph involved a direct translation of the archive's data and data structure into basic RDF, thereby lacking a meaningful, interoperable, and reusable ontology. Furthermore, this approach falls short of satisfying the requirements articulated by the performing arts community. For instance, it is missing a clear distinction between an individual performance event and the original literary work the performance is based on.

The goal is to create and publish the Linked Stage Graph as an open research resource for performing arts. This resource will reuse open standards and ontologies, enabling both an efficient and scientifically accurate exploration of historical performing arts data, as well as easy interconnection with datasets from various sources stored in archives, museums, theatre institutions, and universities. So far, no commonly accepted ontology for historical performing arts data exists. This position paper contributes requirements for data representation and exploration and lessons learned from the ongoing process of creating an application-driven and efficient data model for Linked Stage Graph and more general the performing arts community. The contribution further includes a discussion of the Swiss Performing Arts (SPA) data model as the most fitting and best documented data model and a presentation of modeling adaptations, thereby leveraging existing standards and ontologies like CIDOC-CRM, FRBR, and FRBRoo. The ongoing development process of Linked Stage Graph is documented on GitHub<sup>3</sup>.

Section 2 presents related efforts in the performing arts domain, followed by a discussion of requirements in section 3. The performing arts data model for Linked Stage Graph is discussed in section 4, followed by a conclusion and outlook in section 5.

## 2. Related Work

There are a number of systems, platforms and data models which provide access to (historical) performing arts data. Often, their entry point is focused on either a biographical approach highlighting a relevant person in the performing arts, a regional approach with focus on data created within a city or country, and an institutional approach with data sets stored in a specific archive or theatre. Linked Stage Graph is based on historical data created at Stuttgart State Theatres and provided by the State Archives of Baden-Württemberg, Germany.

Concerning the biographical approaches, there are three platforms to highlight. Ipsen Stage<sup>4</sup>, a performance database about Henrik Ibsen, the Pina Bausch Archive [2] containing AV-material,

---

<sup>1</sup><https://slod.fiz-karlsruhe.de/sparql>

<sup>2</sup><https://slod.fiz-karlsruhe.de/vikus>

<sup>3</sup><https://github.com/ISE-FIZKarlsruhe/LinkedStageGraph/tree/master/LinkedStageGraph2.0>

<sup>4</sup><https://ibsenstage.hf.uio.no/>

program flyers, and costumes as well as Staging Beckett [3], a database of Samuel Beckett plays. Furthermore, Re-Collecting Theatre History [4] provides a data capturing tool and an exploration environment for theatre sciences. Noteworthy approaches with an institutional focus include the Abby Theater Platform [5] and the Specialised Information Service for Performing Arts [6]. Important regional approaches include the Dutch project ONSTAGE [7], the Australian project AusStage [8] and the Swiss Performing Arts (SPA) Platform [9].

The listed projects and platforms are valuable and relevant research resources in the performing arts. However, to the best of our current knowledge, there are no public SPARQL endpoints available for any of them. Moreover, even in cases where ontologies were designed to represent the data, they have not been made publicly available for reuse (e.g. [2]). The SPA project is highly relevant to the development of the Linked Stage Graph, as it provides a comprehensive and well-documented data model. Further details about this data model will be discussed in section 4. Linked Stage Graph is accessible by means of a SPARQL endpoint and all data and development progress is open and documented on Github.

### 3. Requirements for Data Representation and Exploration

To provide Linked Stage Graph as a useful research resource, a requirement analysis has been carried out. The listed requirements were extracted from scientific literature in the domain of the performing arts, as well as workshops with domain experts<sup>5</sup>. The requirement analysis is an ongoing process and will be extended as part of the iterative ontology development process.

- REQ1: **Context.** For each entity in the data collection, provide as much context as possible. No entity in the performing arts exists on its own (e.g. archival object, performance, person, role, stage element) and should be viewed in its context to other entities.
- REQ2: **Perspective.** Provide various perspectives on performing arts data for data exploration to enable a holistic view.
- REQ3: **Interoperability.** Enable the interconnection between disciplines, data sets, archives, performing arts institutes as well as regional and international efforts.
- REQ4: **Persons and Functions.** All persons on, behind and in front of the stage of a performance and their roles and functions are relevant for research.
- REQ5: **Change.** Performing arts are dynamic and the change over time should be represented in terms of persons, occupations, stage design, etc.
- REQ6: **Events.** Performing arts data is often event-based. It should be distinguished between an original work, a production and the performance as an event.
- REQ7: **Stage Elements.** Objects on stage should be captured. If possible, the meaning of an object on stage should be represented (e.g. a chair as an object is used as a throne).
- REQ8: **Querying.** A data model that represents performing arts data should be as lightweight as possible to enable intuitive querying.
- REQ9: **Provenance.** It has to be possible to verify and track research results, e.g. biographical data has to be linked to their data source.

---

<sup>5</sup><https://github.com/ISE-FIZKarlsruhe/LinkedStageGraph/tree/master/LinkedStageGraph2.0>

REQ10: **Data Quality.** The quality of the data used in performing arts research has to be clear and should be quantifiable.

Throughout the process of creating a data model and implementing means of data exploration for the Linked Stage Graph, these requirements are being carefully taken into account. However, not all requirements can be fully met due to the sparsity of the available metadata. This position paper focuses on modeling performing arts data. Based on the requirements above, competency questions have been developed and published on GitHub to aid the modeling process and enable a systematic ontology evaluation.

## 4. Modeling Historical Performing Arts Data on the Example of Linked Stage Graph

This section discusses the main focus of the paper, a data model for the performing arts on the use case of the Linked Stage Graph dataset. Section 4.1 introduces the dataset and a few challenges within. Section 4.3 then provides a modeling based on the Swiss Performing Arts data model along with an adaption for a more simplified lightweight model.

### 4.1. Dataset and Challenges

Linked Stage Graph is a Knowledge Graph created with data provided by the State Archives of Baden-Württemberg, Germany. It contains metadata about performances at the Stuttgart State Theatres from the 1890s to the 1940s. Performances include plays, operas, and ballets. The dataset furthermore contains 7.000 black and white photographs depicting these stage performances, stage design, costume design and rehearsal sessions. The performance types include premieres, new productions, repertoires, and first performances. Shakespeare's Twelfth Night<sup>6</sup> (in German "*Was ihr Wollt*") represents one of the performances. The play ("*Schauspiel*") is a new performance ("*Neuinszenierung*") and was performed on March 11, 1923. Furthermore, information provided by the archive about the performance and production is given in a semi structured manner. For instance, the description lists persons like the stage director ("*Inszenierung*"), stage designer ("*Bühnenbild*") and costume designer ("*Kostüme*"). This performance is associated with three photographs and further metadata including archival identifiers and structures. This performance event is representative for all performances in the dataset considering the representation and richness of the available metadata.

This dataset lacks a coherent and expressive data model that allows to query and explore these historical theatre data efficiently and in a way useful for research purposes (cf. 3). 1) The KG currently lacks interoperability (REQ3) which complicates efforts to connect the dataset with others and the ability to provide more context to the data, e.g. provide information about historical events during a period of a theatre production (REQ1). 2) The original literary works (e.g. "Twelfth Night" by William Shakespeare) are blended together with specific performance events (e.g. the play "*Was ihr Wollt*" on March 11, 1923) into one performance entity. In this example<sup>7</sup>,

---

<sup>6</sup><https://slod.fiz-karlsruhe.de/labw-2-2599390.html>

<sup>7</sup><https://slod.fiz-karlsruhe.de/labw-2-2599390.html>

William Shakespeare is erroneously added to the performance event as *schema:contributor*. This introduces ambiguity and does not allow to clearly extract which metadata concern the original work or the individual performance event. 3) The semi structured data listed in the description were automatically extracted and integrated as entities in the dataset by using too general and unfitting properties. 4) The dataset furthermore contains a large number of blank nodes which significantly increases query complexity and limits exploration (REQ8).

In the following a strategy for Linked Stage Graph and its data model is presented and lessons learned are being discussed.

## 4.2. Linked Stage Graph Modules

The Linked Stage Graph data model is being created in a modular approach:

1. **Performing Arts:** An ontology to represent all data within Linked Stage Graph related to the performances, original works, persons and their functions and places. This is the main focus of this paper and is discussed below.
2. **Archival Structure:** Represents the data structure as provided by the archive. It is based on the (physical) folders in the archive before digitization, includes natural language descriptions of the archival objects and is not intuitive for web-based exploration. However, this archival structure is essential to preserve (REQ9). To represent this structure, RiC-O is being utilized. This module is work in progress and not in the scope of this paper.
3. **Data Analysis:** The photographs in this archival collection contain relevant information about the historical performances. According to REQ7, metadata about stage elements like objects and animals are highly relevant for research purposes. Experiments on analyzing historical theatre photographs have been conducted and a data model which utilizes the Web Annotation Ontology has been developed to integrate results into the KG meaningfully. The results were discussed in [10].

## 4.3. Leveraging Existing Ontologies for Historical Performing Arts Data

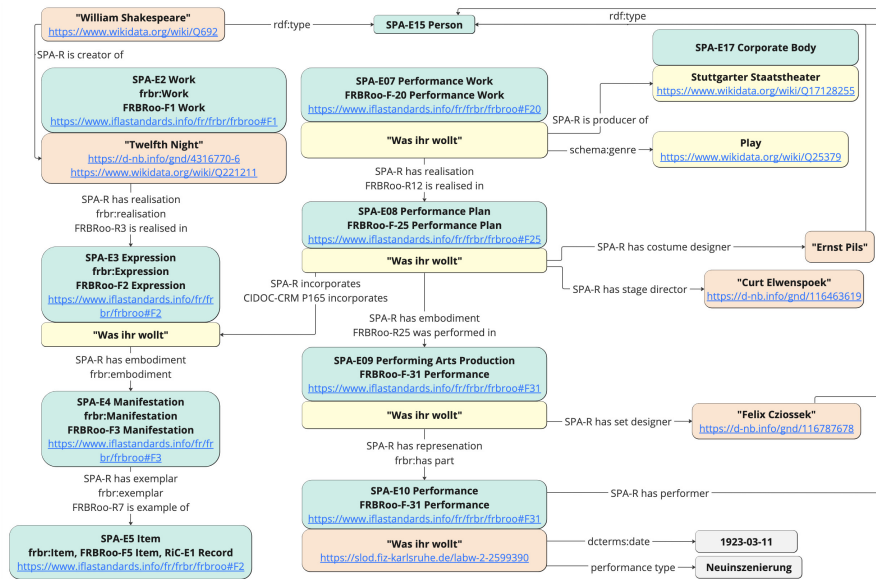
The metadata within Linked Stage Graph (as provided by the archive) blend all information about a single performance event, the original work the performance is based on and the archival objects (e.g. photographs) together. Within theatre history research this is problematic, since a performance may significantly differ from its original work and is a creative work on its own. It is essential to unambiguously model that Shakespeare authored "The Twelfth Night" as a literary source but did not provide creative input to the German play "Was ihr Wollt" as performed on March 11, 1923. To the best of our knowledge, the Swiss Performing Arts (SPA) data model is the best documented data model that is available in the performing arts domain [9]. It is event based and emphasizes the importance of differentiating the original work and the performance event itself (REQ6) by aligning with CIDOC-CRM<sup>8</sup>, FRBR<sup>9</sup> and FRBROO<sup>10</sup>, which is also recommended in a detailed study on German performing arts data in [11]. However, the

---

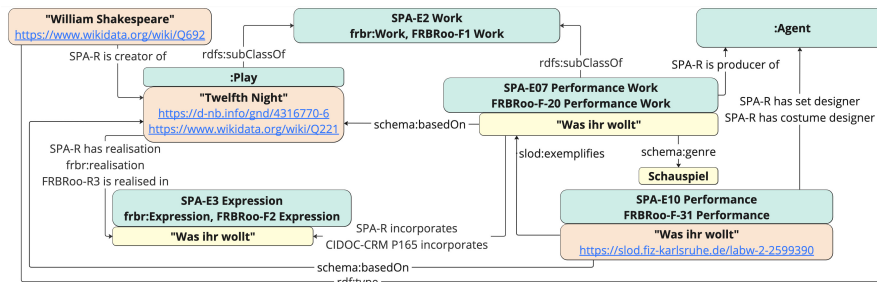
<sup>8</sup><https://www.cidoc-crm.org/>

<sup>9</sup><https://vocab.org/frbr/core>

<sup>10</sup><https://www.iflstandards.info/fr/frbr/frbroo>



**Figure 1:** Leveraging the Swiss Performing Arts Data Model for Linked Stage Graph on the example of the performance "Twelfth Night". Green boxes represent classes, red boxes instances already present in the KG, yellow boxes are instances which have to be newly created to fit the model.



**Figure 2:** Simplified Modeling of "Twelfth Night" utilizing the SPA model

SPA model is not intended for historical performing arts data specifically which often contains sparse metadata, is heterogeneous and depends on varying digitization and storing techniques.

In order to leverage the expressivity and feasibility of the SPA model for Linked Stage Graph (and more general historical performing arts data) performance events as represented in Linked Stage Graph were modeled to fit the SPA and will be discussed in the following. Fig. 1 visualizes a modeling example of William Shakespeare's "The Twelfth Night". All data which regards the original work is modeled using the FRBR model by differentiating between *Work*, *Expression*, *Manifestation*, and *Item*. In this case *Expression* refers to a German translation of the work titled "Was ihr wollt". Unfortunately, it is not possible to provide more information on the translation itself. All data which regards the play as it was performed in Stuttgart in 1923 is represented using FRBRoo. The performance event which took place on March 11, 1923 in Stuttgart is a FRBRoo Performance (and subClassOf CIDOC-CRM E7 Activity). In the current Linked Stage Graph dataset, all metadata are connected to this single entity<sup>11</sup>. However, within the SPA

<sup>11</sup><https://slod.fiz-karlsruhe.de/labw-2-2599390.html>

model there is a differentiation between the *Production SPA-E09* (which may contain a number of performance events and is attributed with a set designer, stage assistant etc.), the *Performance Plan SPA-E08* (which contains directions considering choreographs, and costume design) and the *Performance Work SPA-E07* (the distinct performing arts creation at a conceptual level which is attributed with the production company).

While this detailed differentiation is a real-world representation of the entities, it is highly complex and assumes a rich set of metadata. Currently, only the instance for the performance event (SPA-E10) exists within Linked Stage Graph, i.e. instances for the production, the performance plan and performance work are unknown and would need to be newly created. However, due to the sparse metadata this only increases the number of entities for the same event and query complexity without providing additional useful knowledge. Furthermore, historical theatre data often doesn't provide these detailed information of persons and their functions. A user interested in all persons related to one performance event as depicted in Fig. 1 would be required to query all abstraction levels. Furthermore, a query to list all performance events which are based on a certain *FRBR:Work* would be immensely complex and does not meet REQ8.

Fig. 2 provides a modeling suggestion, which utilizes essential parts of the SPA model but in a more lightweight simplified version to improve query efficiency and exploration capability (REQ8). Most significantly, it extends SPA data model by providing a set of direct relations between entities of different granularities. Thus, such modeling allows for omitting *SPA-08 Performance Plan* and *SPA-09 Performing Arts Production*, by connecting an individual performance "Was ihr wollt" directly to the performance work "Was ihr wollt" via property *slod:exemplifies*, and to the original Shakespeare's work "Twelfth Night" via property *schema:basedOn* (REQ6). This is rational because all persons working on a specific production plan (e.g. costume designer) or production (e.g. set designer) can also be associated with the individual performance events (REQ4). A common research question raised by domain experts is a relation between an original work and its theatrical interpretation. In the original SPA model (Fig. 1), such queries can only be achieved via property path, e.g. `?performanceWork frbroo:R12 [ crm:P165 [ frbr:realisationOf ?originalWork ] ]`. To enable direct and intuitive querying (REQ8) for domain experts the *schema:basedOn* relation is introduced and mapped to a corresponding property path. In Linked Stage Graph, each performance is connected to its Wikidata entity, which can be used to extract the respective genre, e.g. play (REQ1). The genre provided by the archive for each performance is also preserved and attributed to the performance work. This allows to differentiate between the genre of the original literary work and the genre of the performance it is based on.

While the SPA is an extensive and detailed model for performing arts data, it is also complex and assumes a richness in metadata about persons and their functions which often historical cultural heritage collections cannot deliver. Furthermore, the take-up of the SPA model within the performing arts community is rather low. This section provided a simplified version of the SPA model. While reusing open standards and ontologies like CIDOC, FRBR, and FRBRoo, the more lightweight model presented in Fig. 2 is interoperable (REQ3) and easy to further extend and specify. The data within Linked Stage Graph is with respect to its richness and its information structure a representative resource for historical archival and performing arts data in Germany [12] and the presented efforts are generalizable beyond the presented dataset.

## 5. Conclusion and Outlook

This position paper reports on the journey towards an expressive and efficient data model for Linked Stage Graph and more general the performing arts community by leveraging existing standards and ontologies like CIDOC-CRM, FRBR, and FRBRoo. To date, there is no widely accepted ontology for historical performing arts data. Nevertheless, this paper also demonstrates that existing ontologies can be used and extended to render historical performing arts data interoperable, efficient to query and easy to explore. In a next step, the discussed modeling adaptations will be evaluated and integrated into Linked Stage Graph with the goal to publish it as an open, interconnected and scientifically accurate performing arts research resource.

## References

- [1] T. Tietz, J. Waitelonis, K. Zhou, P. Felgentreff, N. Meyer, A. Weber, H. Sack, Linked stage graph., in: SEMANTICS Posters&Demos, 2019.
- [2] B. Thull, K. Diwisch, V. Marz, Linked data im digitalen tanzarchiv der pina bausch foundation, in: Corporate Semantic Web: Wie semantische Anwendungen in Unternehmen Nutzen stiften, 2015, pp. 259–275.
- [3] A. McMullan, T. McTighe, D. Pattie, D. Tucker, Staging beckett: constructing histories of performance, *Journal of Beckett Studies* 23 (2014) 11–33.
- [4] N. Probst, V. Pinto, Re-collecting theatre history, theaterhistoriografische nachlassforschung mit verfahren der digital humanities, in: B. Wihstutz, B. Hoesch (Eds.), *Neue Methoden der Theaterwissenschaft*, transcript, 2020, pp. 157–179.
- [5] M. Bradley, A. Keane, The abbey theatre digitization project in nui galway, *New Review of Information Networking* 20 (2015) 35–47.
- [6] J. Beck, M. Büchner, S. Bartholmei, M. Knepper, Performing entity facts: The specialised information service performing arts, *Datenbank-Spektrum* 17 (2017) 47–52.
- [7] F. R. E. Blom, H. Nijboer, R. van der Zalm, ONSTAGE, the Online Data System of Theatre in Amsterdam from the Golden Age to Today, *Research Data Journal for the Humanities and Social Sciences* 5 (2020) 27–40.
- [8] J. Bollen, Data models for theatre research: People, places, and performance, *Theatre Journal* (2016) 615–632.
- [9] B. Estermann, F. Julien, A linked digital future for the performing arts: Leveraging synergies along the value chain, in: Canadian Arts Presenting Association (CAPACOA) in cooperation with the Bern University of Applied Sciences, 2019.
- [10] T. Tietz, J. Waitelonis, M. Alam, H. Sack, Knowledge graph based analysis and exploration of historical theatre photographs., in: *Qurator*, 2020.
- [11] W. Schneider, H. Fülle, C. Henniger, *Performing the archive: Studie zur Entwicklung eines Archivs des Freien Theaters*, volume 34, Georg Olms Verlag, 2018.
- [12] F. Hoppe, T. Tietz, D. Dessi, N. Meyer, M. Sprau, M. Alam, S. Sack, Harald, et al., The challenges of german archival document categorization on insufficient labeled data, in: *CEUR Workshop Proceedings*, volume 2695, CEUR-WS, 2020, pp. 15–20.