

# Ontologien - Datenmodelle - Mapping: Theadok.at

Datenmodell, Modellierung und Beispiele

Creators:

Klaus Illmayer

Sara Tiefenbacher

Context:

Präsentation beim Treffen der *Unterarbeitsgruppe Ontologien* der [Arbeitsgruppe Archiv](#) der [Gesellschaft für Theaterwissenschaft](#) (Online, 18.09.2020)

License:

Diese Slides sind lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](#).

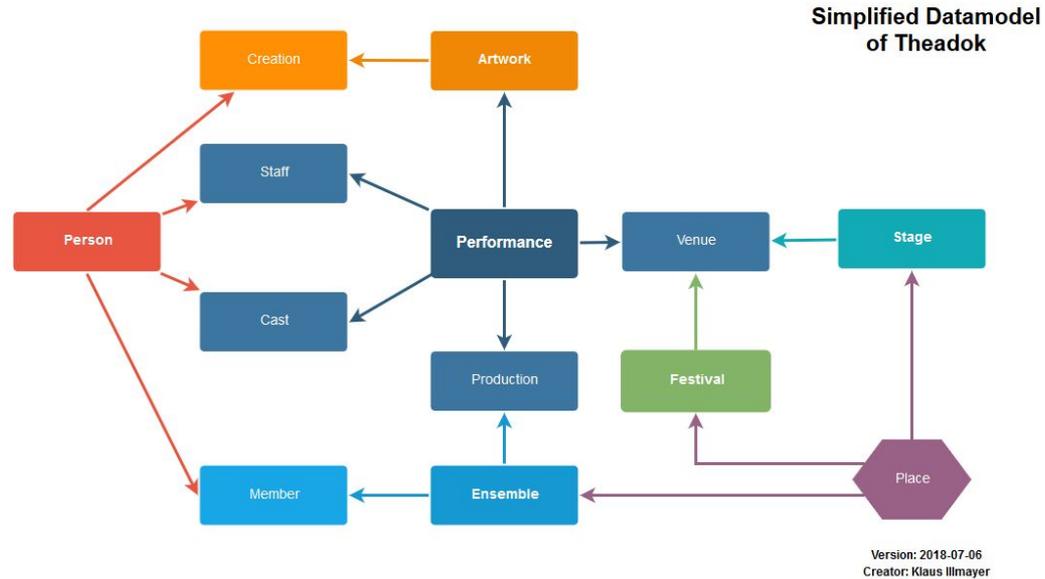
# Datenmodelle vs. Ontologien #1

Datenmodell:

- Konkreter Ausdruck wie Daten **strukturiert** (u.a. in Feldern) und in einer Datenbank **gespeichert** werden (alternativ als “data schema” bezeichnet)
- Abhängig von der gewählten **Technologie**:
  - **Individuelle** Strukturen bei Eigenentwicklungen
  - **Vorgegebene** Strukturen bei externen Programmen z.B. CMS wie Drupal, Wordpress
  - Datenbanksysteme können einen Unterschied machen z.B. relational vs. graph database
- Idealerweise gibt es **Vorgaben** z.B. Standards, die umgesetzt werden
  - Tendenz bei existierenden Datenbanken das **Modell weiterzuführen**
    - alternativ: iterative **Weiterentwicklung** oder komplette **Datenmigration**
  - Neue Projekte - die erst Daten generieren - sind freier in ihrer Wahl (hier können Ontologien eine gute Vorlage geben)

# Datenmodell Theadok

<https://theadok.at/about/data>



# Datenmodelle vs. Ontologien #2

Ontologien sind eine **spezifische Form von Datenmodellen** (claim)

- Abstrakte, konzeptuelle Repräsentationen: nicht an konkrete Daten gebunden
- Versuch auf logischer Basis **Beziehungen** zwischen **Entitäten** auszudrücken
- **Entitäten** sind grundlegende, idealerweise nicht teilbare **Basisdatentypen**
- Üblicherweise sind Ontologien “**domain driven**” (Ausnahmen wie CIDOC CRM, die den Anspruch haben, die “Welt zu erklären” bzw. so weit wie möglich in einem Modell abbildbar zu machen => sehr komplexe Strukturen)
- Ontologien werden nicht zwingend in der Praxis (= Datenbanken) verwendet
- Idealerweise “**common sense**” einer Domain für den Austausch von Daten
- Nicht zu verwechseln mit **Dateiformaten** (ebenfalls Austausch von Daten):  
Ontologien geben eine Struktur aber nicht die Ausdrucksform vor

# Datenmodelle vs. Ontologien #3

Zugespitzt formuliert:

- **Datenmodelle** sind **pragmatisch** am technischen Umfeld und an der **Praxis** orientiert
- **Ontologien** sollten **logisch konsistent** (“reasoning”) eine Domain ausdrücken und sind **theoretische Konstrukte**

Darum selten 1:1-Implementierungen von Ontologien in laufenden Systemen aber als **Austauschformat** (via “**mapping**”) oder als Diskussionsformat von Bedeutung

**Gründe** dafür: Ontologien oft überkomplex und umfassend und dadurch technisch aber auch auf Abläufe (z.B. Eingabemasken) bezogen **umständlich umzusetzen**

# Mapping #1

Mapping: Wichtige **digitale Methode** um Daten zwischen Datenmodellen **auszutauschen** oder auch zu **verschränken** (z.B. “enrichment”)

- Ein Mapping erläutert, wie Daten aus einer Datenbank **in Beziehung** gebracht werden können mit Daten aus einer anderen Datenbank
- **Vollständiges Mapping** meist nicht möglich
  - Entweder existieren im Ziel-/Quelldatenmodell eines oder mehrere Datenfelder nicht
  - oder die Inhalte von Feldern sind nicht ohne Transformation übertragbar (z.B. mehr als eine Information wird in einem Datenfeld gepackt)
  - oder scheinbar ähnliche Felder sind es nicht
  - oder es gibt im Zieldatenmodell eine strikte Vorgabe von Inhalten (z.B. ein kontrolliertes Vokabular) die im Quelldatenmodell anders oder laxer behandelt werden

# Mapping #2

**Proposal:** Eine **bottom-up Ontologie-Entwicklung** würde von **bestehenden Datenmodellen** ausgehen und versuchen, diese auf **konzeptueller Ebene in Einklang** zu bringen:

- **Identifikation** von Übereinstimmungen, Lücken, kreativen Modellierungen
- Gründe für **Unterschiede** nachforschen: kann mit Untersuchungsgegenstand zusammenhängen oder auch konzeptuelle Fehler sein oder auf vergangene Entscheidungen beruhen (z.B. Einsatz einer spezifischen Software)
- Sammlung und **Dokumentation** von **Varianten**
- **Erweiterung existierender Ontologien** (z.B. [SPA](#)) sollte bevorzugt werden

# Mapping #3

- **Kompromisse** wichtig aber zu viel Rücksichtnahme problematisch
- **Bestandsaufnahme** über gemeinsame Analyse von **Datenmodellen** und den **Daten** selbst
- [FID](#) hat die beste Erfahrung
- **Mapping** als Methode um **Gemeinsamkeiten und Problemfelder** zu identifizieren sowie einen **Überblick** zu den **Interpretationen** zu erhalten
- Alleiniges Mapping (ohne Datenmanipulation bei der Quelle) oft mit Qualitätsverlust verbunden - deswegen **Datenmanipulation** mitbedenken
- Ohne Datenmanipulation (bzw. “data processing”) ist ein Mapping unzureichend => Mapping ohne Programmierkenntnisse machbar, Datenmanipulation eher nicht

# Überlegungen zu Mapping/Ontologien #1

- Nicht immer notwendig, manchmal reicht ein **simpler Link** (z.B. zu Archivmaterial): **wo endet die Domain?**
- Upgrade existierender Datenbanken nicht ohne weiteres möglich
- Entwicklung neuer Datenbanken sollte **Erfahrungen/Vorschläge berücksichtigen** (passiert eher selten)
- Idee der “**Annäherung**” (convergence): wenn “best practices” existieren und kommuniziert werden, dann orientieren sich andere daran
- Grundsätzlich: statt einer Ontologie (die sich an Vollständigkeit orientiert) kann auch ein **minimales Datenschema** (enthält nur Basisinformationen, auf die sich die meisten einigen können) ein Ziel sein (bzw. als eine Vorstufe zu einer Ontologie)

# Überlegungen zu Mapping/Ontologien #2

- Einbindung von **Technikexpertise** ist wichtig bei Ontologieentwicklung
- Gilt auch für **Implementierung**: Wird eine Ontologie nicht verwendet (= implementiert), dann ist sie wenig sinnvoll
- Zu berücksichtigen sind auch **Fallbeispiele, Analysen, Visualisierung** die gewünscht werden, sonst gibt es Enttäuschung
- **Ontologien sind Mittel zum Zweck** nicht die Lösung: Entwicklung und Pflege von Software sowie Eingabe von Daten wird damit nicht automatisch vereinfacht stattdessen wird auf **Interoperabilität** abgezielt, die sich meist erst später rentiert u.U. erst in anderen Projekten (**Sharing, Collaboration, Openness** muss akzeptiert werden - große soziale Herausforderungen z.B. Konkurrenz bei Förderungen)

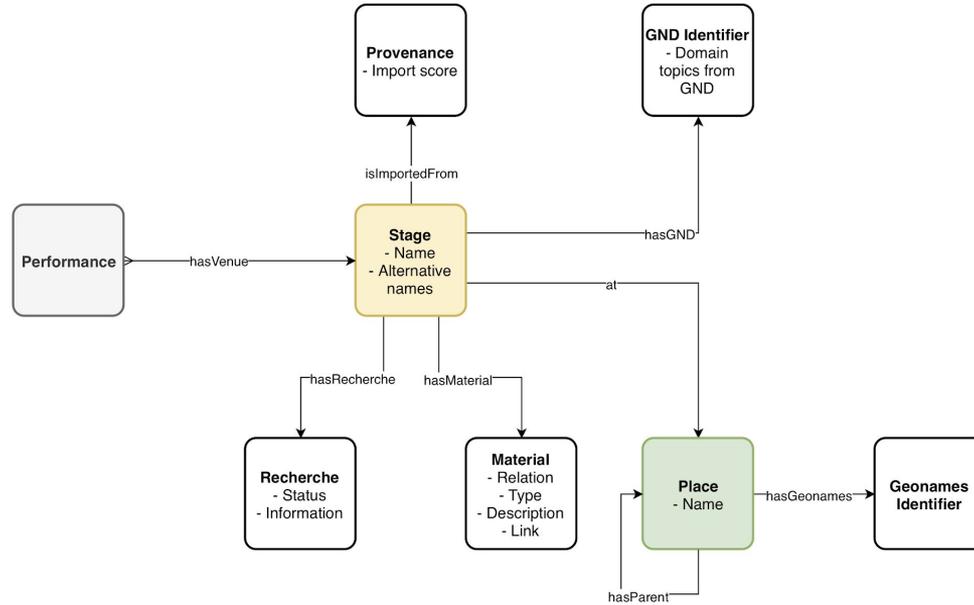
# Vorschlag für Arbeitsprogramm UAG Ontologie

- Sammeln von Beispielen zu Datenmodellen und Daten
- Gemeinsames Mapping zu SPA um Stärken/Schwächen zu identifizieren
- Spezialfälle berücksichtigen z.B. Ausfälle von Inszenierungen
- Dokumentation dieser Mappingpfade
  - Kontexte darlegen, in denen Daten produziert werden
  - Fachliche Perspektiven einbringen: von wo kommen Daten, wie können diese Daten in theaterwissenschaftliche Auseinandersetzung eingebracht werden
  - Generierung von Daten im Fach selbst
  - Analyseversuche und -absichten darlegen
- Zunächst Verbesserung von SPA und aus den Erfahrungen daraus weitere Rückschlüsse ziehen
- Festlegen wo die Domain der UAG Ontologie beginnt und endet

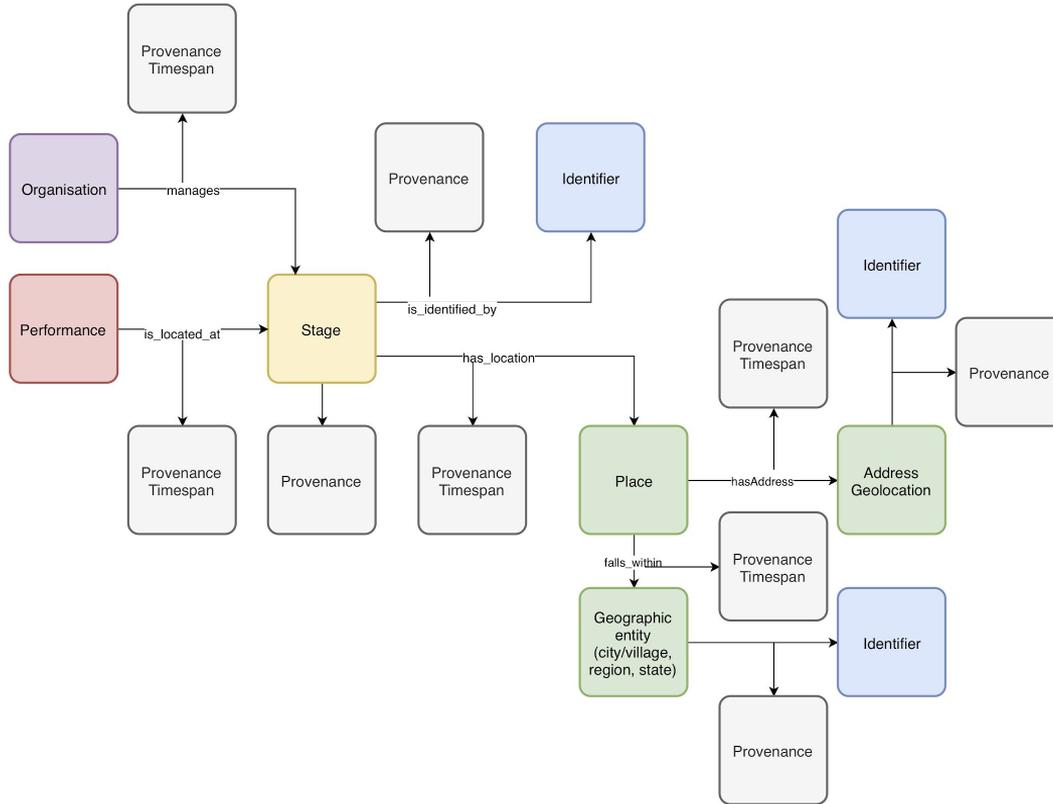
# Datenmodell Theadok - Mapping zu Datenmodellen/Ontologien

# Datenmodell Theadok

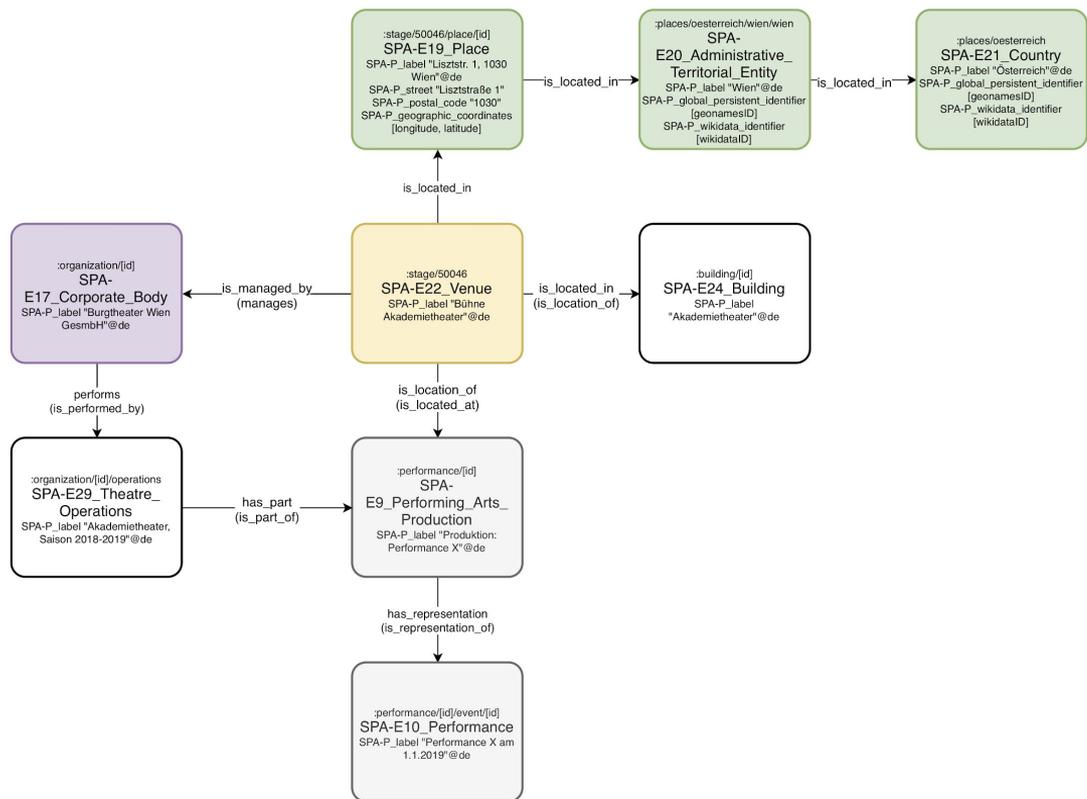
- Hat eine **Geschichte**, die sich im Datenmodell widerspiegelt
- Keine ideale Lösung
- **Erweiterung** des Modells um **Datenqualität** zu erhöhen: Granulation von Daten, mehr Detaildichte, neue Informationen hinzufügen (**mehr Kontext**), bessere und genauere Verbindung von Entitäten und Einfügen von **Andockstellen** zu externen Datenbanken (meist “Identifizier”)
- **Iteration ist Herausforderung**: Umstrukturierung bedeutet ein Mapping zur Datenmodellerweiterung und inkludiert Datenanalyse, Datenmanipulation, oft manuelle Anpassungen und Tests
- **Lokales Enrichment** von Daten durch Vernetzung zu anderen Datenbanken
- Vernetzung mit weiteren Daten hat Priorität: mehr Dichte erzeugen



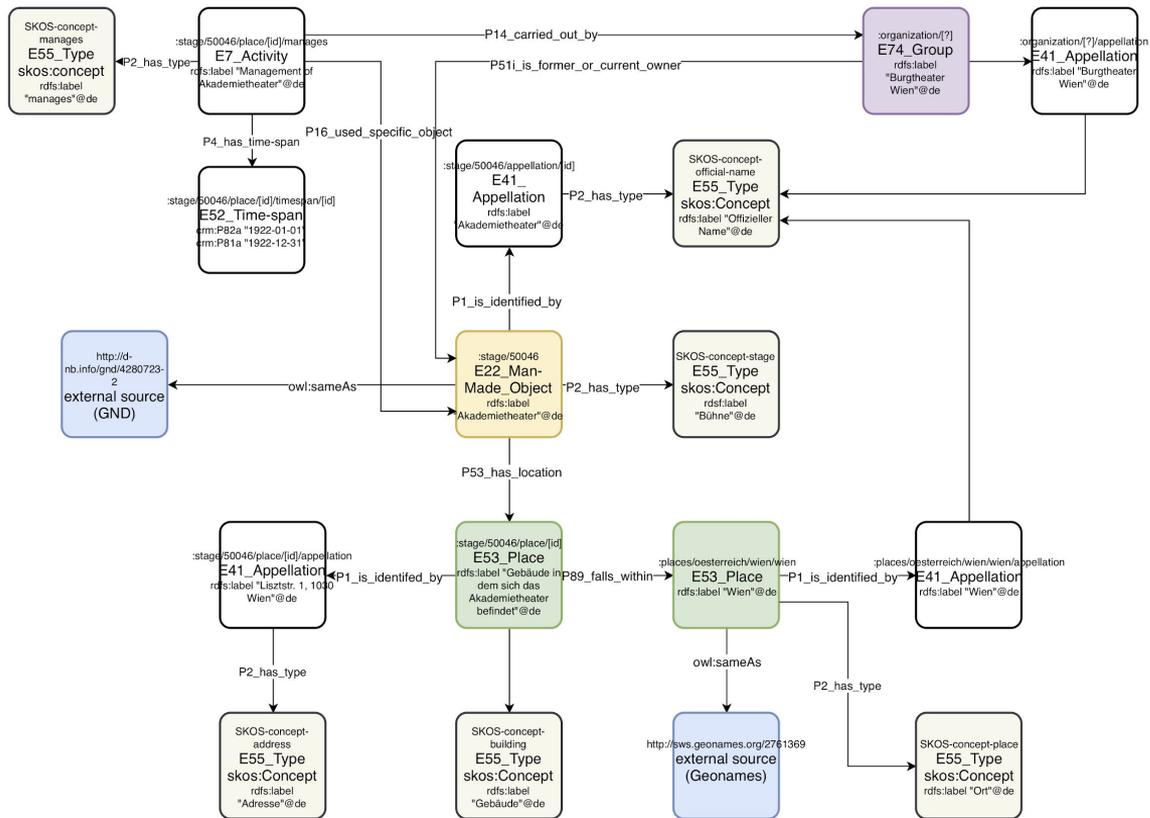
Venue Theadok 1: current state



Venue Theadok 2: current iteration



Venue SPA: mapping proposal



Venue Theadok: CIDOC mapping

Vernetzung von Daten in Theadok  
(manchmal Link, manchmal Mapping)

# Datenmodellierung - Eintrag zu Inszenierung

Beispiel Inszenierungseintrag: "Was geschah, ..." (1992); Theadok identifier: <https://theadok.at/performance/5259>

Suche nach Inszenierungs- oder Stücktitel, Personen-, Bühnen- oder Ensemblename

## Was geschah, nachdem Nora ihren Mann verlassen hatte oder Stützen der Gesellschaft

Identifier: <https://theadok.at/performance/5259> (Last change: 29.12.2017-13:35)

Inszenierung 

**Allgemeines**

Aufführungssprache	Deutsch
Sparte	Sprechtheater

**Relationen**

hat Textvorlage	Was geschah, nachdem Nora ihren Mann verlassen hatte oder Stützen der Gesellschaft
Collections	Theadok Import

**Produktionskontext**

Produktionsort	Wien
----------------	------

**Inszenierungskontext**

Aufführung	Neuinszenierung
Premiere	So., 08.03.1992 91/92
	Volkstheater «Wien»

**Personen**

Regie	Werner, Emmy
Ausstattung	Zuzanek, Mimi
Choreographie	Modra, Blanka
Regieassistenz	Spannbauer, Margarete

**Besetzung**

Nora	Eckert, Andrea
------	----------------

# Datenmodellierung - Eintrag zu Inszenierung (2. Teil)

Inszenierungskontext	
Aufführung	Neuinszenierung
Premiere	So., 08.03.1992 91/92
	Volkstheater <Wien>

Personen	
Regie	Werner, Emmy
Ausstattung	Zuzanek, Mimi
Choreographie	Modra, Blanka
Regieassistentz	Spannbauer, Margarete

Besetzung	
Nora	Eckert, Andrea
Arbeiterinnen	Lippert, Cornelia
Vorarbeiter	Hall, Bernhard
Sekretärin	Mertinz, Johanna
Arbeiterinnen	Schuchter, Gabriele
Arbeiterinnen	Eckert, Andrea
Nora	Lippert, Cornelia
Personalchef	Obonya, Cornelius
Die Arbeiterin Eva	Schell, Hertha
Arbeiterinnen	Arens, Babett
Nora	Arens, Babett
Nora	Schuchter, Gabriele
Nora	Frey, Beatrice
Arbeiterinnen	Frey, Beatrice

Materialien	
Quelle ist	Programmzettel
INSZNR auf CD-ROM ist	1507699

# 1. Datenvernetzung: Theadok und Archivmaterial

- Inszenierungsdatenbank Theadok (<https://theadok.at>)
- Archiv und theaterhistorische Sammlung  
(<https://fm.univie.ac.at/sammlungen-einrichtungen/archiv-und-sammlungen/>  
und Archiv-Datenbank (<https://archiv-fm.univie.ac.at/>)

# Beispiel 1: Archivmaterial

Inszenierung “Alles für Marion” (1932)

<https://theadok.at/performance/130850>

Verweis auf Archivmaterial: 2 ‘graphic records’ (Aufführungsdokumentation) in Foto- und Graphiksammlung / Bestand des Archivs und der theaterhistorischen Sammlung am tfm

<https://archiv-tfm.univie.ac.at/record-set/676>

# Beispiel 2: Rezensionen/Theaterkritiken

Inszenierung “Der Bauer als Millionär” (1978)

<https://theadok.at/performance/22364>

Verweis auf Rezensionen im Theadok Archiv (interner Zugang, liegen digital vor)

Signatur: Theadok Archiv 36/2147-2152

## 2. Datenvernetzung Theadok und AV-Materialien

- Inszenierungsdatenbank Theadok (<https://theadok.at>)
- Videothek des tfm Instituts  
(<https://tfm.univie.ac.at/standorte-oeffnungszeiten/videothek/>)

Theadok ist derzeit in der Beta-Version. Das bedeutet, dass sowohl bei den Funktionen, der Darstellung als auch den Daten noch Anpassungen nötig sind. Für Hinweise auf fehlerhafte Daten oder für Vorschläge um Theadok besser nutzbar zu machen, freuen wir uns über eine Nachricht an [theadok.tfm@univie.ac.at](mailto:theadok.tfm@univie.ac.at)

Suche nach Inszenierungs- oder Stücktitel, Personen-, Bühnen- oder Ensemblename

SEARCH

# Leonce und Lena

Identifier: <https://theadok.at/performance/8889> (Last change: 27.05.2019 - 14:35)

## Inszenierung

### Allgemeines

Sparte [Sprechtheater](#)

### Relationen

hat Textvorlage [Leonce und Lena](#)

Collections [Theadok Import](#)

### Produktionskontext

Produktionsort [Salzburg](#)

### Inszenierungskontext

Aufführung [Neuinszenierung](#)

Premiere [Di, 12.08.1975](#)

# Beispiel

Inszenierungseintrag in  
Theadok:

<https://theadok.at/performance/8889>

# Beispiel

Aufnahme von Links zu  
A/V-Material

Verweist unter anderem auf  
eine Ausstrahlung in ORF1  
und auf ein Video,  
veröffentlicht von  
Arthaus-Musik

Es fehlt noch der Link zum  
Eintrag in der Videothek tfm  
Wien (Authentifizierung)

Produktionskontext	
Produktionsort	<a href="#">Salzburg</a>

Inszenierungskontext	
Aufführung	<a href="#">Neuinszenierung</a>
Premiere	 Di, 12.08.1975 74/75
	<a href="#">Landestheater &lt;Salzburg&gt;</a> <a href="#">Salzburger Festspiele</a>

Personen	
Regie	<a href="#">Schaaf, Johannes</a>
Bühnenbild	<a href="#">Minks, Wilfried</a>
Kostüme	<a href="#">Göbl, Christian</a>

Materialien	
Quelle ist	<a href="#">Pressekritik</a>
	<a href="#">Theadok Archiv</a> 8-1311-133
INSZNR auf CD-ROM ist	1511309
Aufzeichnung	<a href="#">Ausstrahlung</a>
	<a href="#">ORF1</a> 01.08.1993
Aufzeichnung	<a href="#">Video</a>
	<a href="#">Arthaus-Musik</a> <a href="http://arthaus-musik.com/de/dvd/theater/media/details/Leonce_und_Lena.html">http://arthaus-musik.com/de/dvd/theater/media/details/Leonce_und_Lena.html</a>

# Beispiel: Die Falle - Theadok

## Die Falle

Identifier: <https://theadok.at/performance/5149> (Last change: 29.12.2017 - 13:35)

### Inszenierung

#### Allgemeines

Aufführungssprache [Deutsch](#)

Sparte [Sprechtheater](#)

#### Relationen

hat Textvorlage [Die Falle](#)

Collections [Theadok Import](#)

#### Produktionskontext

Produktionsort [📍 Wien](#)

#### Inszenierungskontext

Aufführung [Neuinszenierung](#)

Premiere [📅 Mi, 05.06.1991  
90/91](#)

Derniere [📅 Fr, 07.06.1991](#)

# Beispiel: Die Falle - Videothek tfm Wien

## Detailanzeige

<< >>

Viedodokument 1 von 1

[Zur Trefferliste](#)

<b>Status</b>	<b>Nicht entlehnt   Nicht vorgemerkt   <a href="#">Bestellen</a></b>
Signatur	G7628
	
Backup des Videos	Sie befinden sich zur Zeit nicht im Datennetz der Uni Wien. Bitte stellen Sie zum Abspielen des Videos eine <a href="#">VPN-Verbindung</a> her. Anschließend aktualisieren Sie bitte diese Seite.  <a href="file:///share.univie.ac.at/tfm-media/video/G/G7000-G7999/G7628/VIDEO_TS/VIDEO_TS.IFO">file:///share.univie.ac.at/tfm-media/video/G/G7000-G7999/G7628/VIDEO_TS/VIDEO_TS.IFO</a> <a href="file:///share.univie.ac.at/tfm-media/video/G/G7000-G7999/G7628/">file:///share.univie.ac.at/tfm-media/video/G/G7000-G7999/G7628/</a>

### Titel

Titel	<b>FALLE</b>
Alternativtitel	Die Falle

### Allgemeine Informationen

Gattung	Sprechtheateraufzeichnung
Produktionsland	AT
Erstaußführungsjahr	1991
Tonspuren	Deutsch (Mono)
Farbe	Farbe
TV-Norm	PAL
(Video)system	DVD/Region 0 (Codefree)
Bildformat	1,33:1
Dauer	103.5 min
TV-Kanal	orf
Schlechte Bildqualität	Ja

### Personen

Anlaß	wiener festwochen 1991
-------	------------------------

### 3. Datenvernetzung: Kontexte - wie modellieren?

Beispiel “Gastspiele”

“Das Leben Josefs”/ “Żywot Józefa” (1965): Gastspiel des Teatr Narodowy / Polnisches Nationaltheater Warschau am Burgtheater Wien

Premiere in Warschau: 3.04.1965; Premiere in Wien: 10.10.1965

<https://theadok.at/performance/21768>

<http://encyklopediateatru.pl/przedstawienie/7781/zywot-jozefa>

Kontext: Wiedereröffnung des „Teatr Wielki” (Großen Theaters) in Warschau => “Żywot Józefa” im Festprogramm “200 Jahre Nationaltheater” November 1965

# Beispiel eines Mappings

# Mapping Bahr-Schnitzler-Inszenierungen #1

Datensatz aus der Literaturwissenschaft basierend auf Edition von Briefwechsel Arthur Schnitzler und Hermann Bahr

- Arthur Schnitzler am ACDH-CH: <https://schnitzler.acdh.oeaw.ac.at/>

Datensatz ist Nebenprodukt dieser Edition und enthält alle erwähnten Aufführungen in diesem Briefwechsel

- Beispiel “Das Märchen”, UA 1893:  
<https://bahrschnitzler.acdh.oeaw.ac.at//index.html?id=T030010,D041094,L041002>

# Mapping Bahr-Schnitzler-Inszenierungen #2

Verknüpfung der verschiedenen Informationen zu Schnitzler über cross-project-Normdatei PMB (Personen der Moderne Basis)

- PMB: <https://pmb.acdh.oeaw.ac.at/>
- Mehrere Entities darunter auch Events:  
<https://pmb.acdh.oeaw.ac.at/apis/entities/entity/event/list/>
- UA “Das Märchen”:  
<https://pmb.acdh.oeaw.ac.at/apis/entities/entity/event/35068/detail>
- API call: <https://pmb.acdh.oeaw.ac.at/entity/35068/> (bietet mehr Information als zur Verfügung gestellter Datensatz im XML-Format)

# Mapping Bahr-Schnitzler-Inszenierungen #3

```
<?xml version="1.0" encoding="UTF-8"?>
<root>
  <row>
    <Heading0>20002</Heading0>
    <Nummer>A020002</Nummer>
    <Titel>Das Märchen</Titel>
    <Erscheinungsdatum>1891</Erscheinungsdatum>
    <Aufführung>UA 1. 12. 1893 Wien, Deutsches Volkstheater</Aufführung>
    <Zyklus/>
    <Autor>A002001</Autor>
    <Autorname>Schnitzler</Autorname>
  </row>
  <row>
    <Heading0>20005</Heading0>
    <Nummer>A020005</Nummer>
    <Titel>Das Tschaperl</Titel>
    <Erscheinungsdatum>1896</Erscheinungsdatum>
    <Aufführung>UA 27. 2. 1897 Wien, Carl-Theater; A 24. 9. 1897 Berlin, Lessing-Theater</Aufführung>
    <Zyklus/>
    <Autor>A002002</Autor>
    <Autorname>Bahr</Autorname>
  </row>
  [...]

```

# Mapping Bahr-Schnitzler-Inszenierungen #4

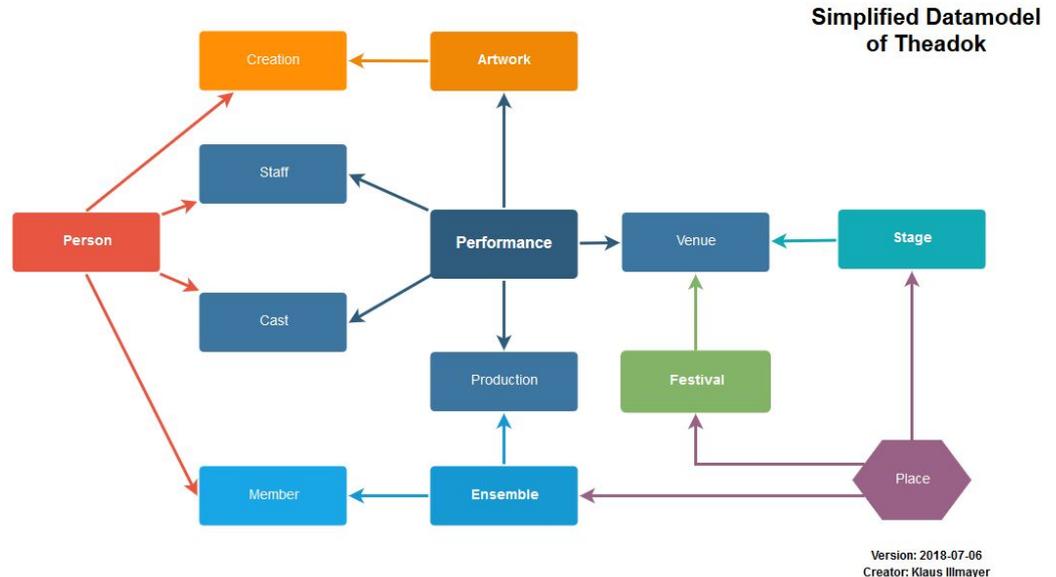
Einfaches Mapping:

- “Titel” => (Art)Work - Performance
- “Erscheinungsdatum” => Production
- “Aufführung” => Production / Stage
- “Autorname” => Person - Creation

Schon mit einigen Schwierigkeiten verbunden

- Selbe Titel für Work/Performance
- Wiederholung Datum
- Aufführung enthält viele unterschiedliche Informationen (“data processing” notwendig)

Wichtig: Ist oberflächliches Mapping,  
entspricht nur grob dem tatsächlichen Modell



# Mapping Bahr-Schnitzler-Inszenierungen #5

- Für ein ausführliches Mapping hier zu “Stage” fehlen Informationen (aber ist auch nicht Absicht des Datensatzes)
- “Titel” => is\_located\_at => “Aufführung” (“Provenance” => “Bahr-Schnitzler-Daten”)

