Connecting archival records with performance identifiers

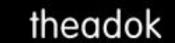
Hyperstage. Global Congress on Digital Innovation in the Performing Arts 04.10.2024

Klaus Illmayer (https://orcid.org/0000-0001-7253-996X)
University of Vienna - Department of Theatre, Film and Media Studies









Overview

- **Performance identifiers** on the example theadok.at
- Theater archives and collections on the example AuS-tfm Vienna
- Connecting information with the help of performance identifiers on the example of the HIC and the HoTS project
- Problems with performance identifiers
- Our current idea for an approach
 - Sketch
 - Challenges
 - Takeaways

Aim of performance identifiers

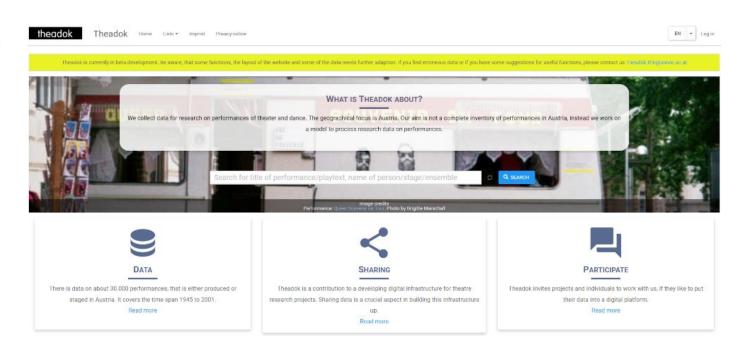
- Performance identifiers a longstanding discussion in digital theatre studies:
 - There is the desire for an authority file of performance identifiers as there are ones for works and persons like GND, VIAF
 - But performances (in general: events) do have a more fuzzy state as works have and there
 is not such a big contributor base as libraries have
- Advantages of (persistent) performance identifiers
 - Claim to speak about the same event
 - All kind of searches especially federated search queries and support of linked open data
 - Connecting with all kind of materials that document or belong to a performance
 - Collection of information from different sources combined to a dense (meta)dataset
 - Allow different research approaches to give input to this collection of information
 - Separate complex information about performance from input forms

Example: theadok.at

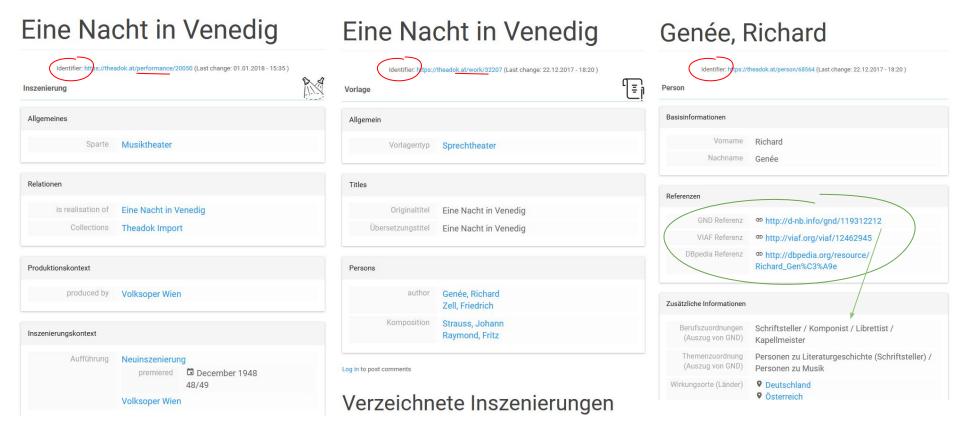
Collection of metadata on theater and dance performances in Austria from 1945 to around 2000 based on announcements, programs, reviews

Lot of efforts to collect the data on the performances but still very different kinds of quality depending on the source

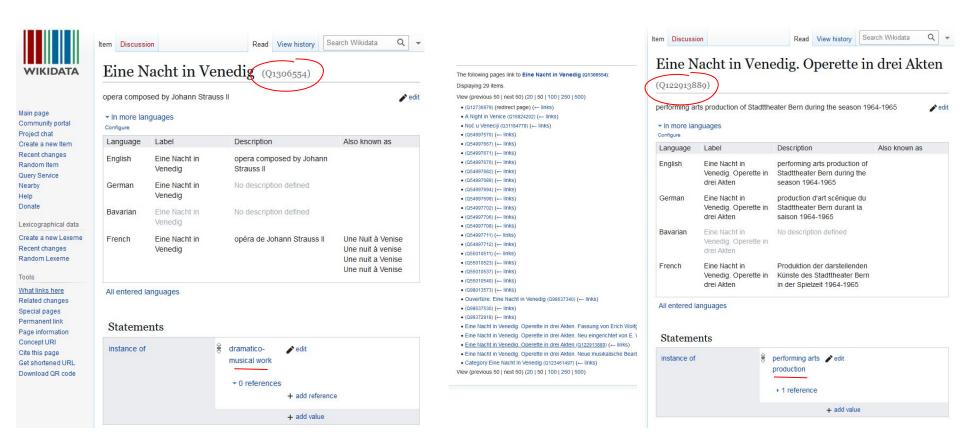
https://theadok.at



Performance identifiers in theadok.at

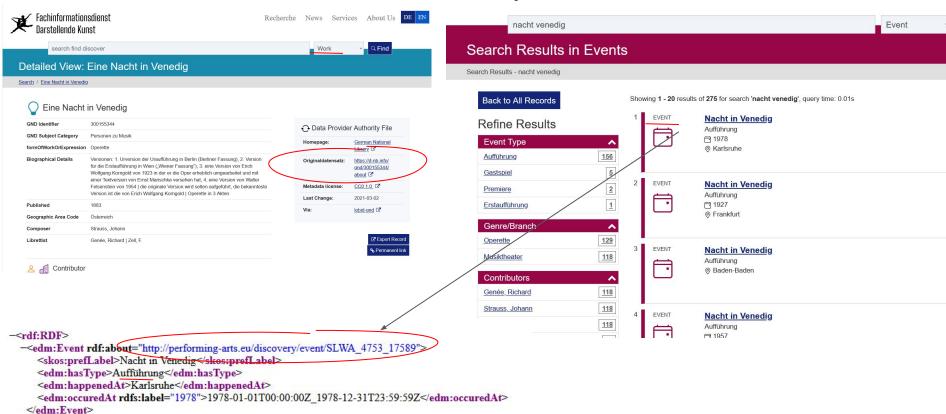


Performance identifiers at other places: Wikidata



Performance identifiers at other places: FID DK

</rdf:RDF>



Handling in (theater) archives

- Performance metadata is often only partially handled by dedicated input fields
- It usually can't deal with ambiguity
- Tendency to record information about persons (directors, actors), companies or theater stages but not performances => organisation of archival records
- Connection of archival objects to performance plan or to one performance attribution of archival objects can be unclear
- Often there is a strong connection with works which may be effective with "traditional" european-style forms of theater but performances are more than this: circus, stand-up comedy, avantgardistic theater forms, amateur theater, etc.
- This all leads to an underperforming handling of metadata on performances and no specific performance identifiers => there is mostly no entity "performance" in an archival database or the corresponding data model

Example: Theatermuseum Austria



Szenenfoto:

Eine Nacht in Venedig

Szene aus "Eine Nacht in Venedig" von C. Walzel und R. Genée, Musik von J. Strauß, Volksoper

Fotograf/Urheber/in: Josef Palffy, Wien

Autor/in: Camillo Walzel (1829 Magdeburg - 1895 Wien) - GND

Autor/in: Richard Genée (1823 Danzig - 1895 Baden bei Wien) - GND

Komponist/in: Johann Strauss (Sohn) (1825 St. Ulrich bei Wien - 1899 Wien) - GND

Regie: Otto Schenk (geb. 1930 Wien) - GND

Datierung: 22.5.1989

Spielstätte: Wien \ Volksoper

Material/Technik: Fotopapier glänzend, S/W

Maße: 12,9 cm × 17,7 cm Inv. Nr.: FS PSE164410

© Theatermuseum, Wien

Permalink (zitierbarer Link) zu dieser Seite: www.theatermuseum.at/de/object/1502897/



AuS-tfm: Archive and theater collections at tfm@Vienna

Established at the founding of the theatre department 1943

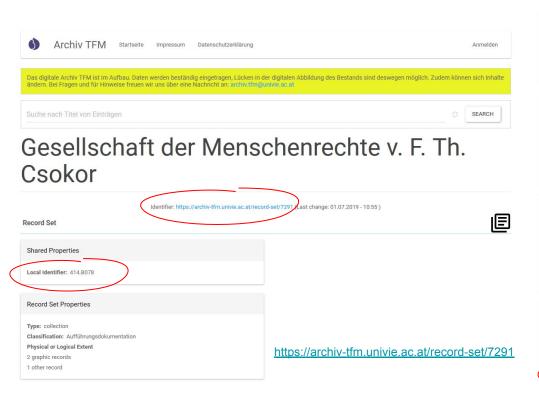
Was **partly lost** but many holdings were found in the last two decades by Birgit Peter, the head of Aus-tfm

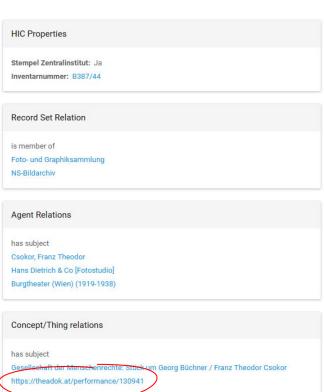
No recordings of the objects where available thus we started a project (HIC) to register parts of the archive in a **digital archive information** system and also debating the national-socialism background

https://archiv-tfm.univie.ac.at/



Identifiers at AuS-tfm

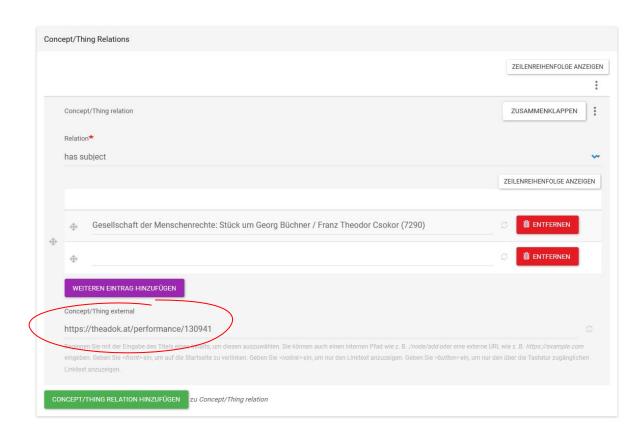




Example: AuS-tfm and theadok.at

using a dedicated archive data model (RiC: Records in Contexts) => does not foresee the connection of performances thus using Concept/Thing relation

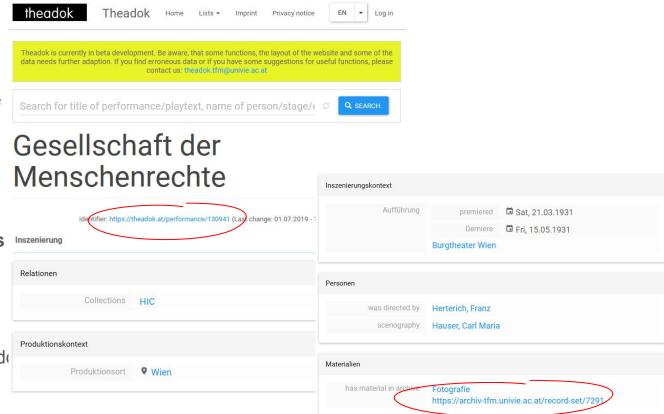
Using a manual copy & paste way to connect data



Connecting archival records with performance identifiers

Example: **HIC** (FWF funded project: Historiography – Ideology – Collection. Research-based Digitizing of Historical Theater Material from the "Zentralinstitut für Theaterwissenschaft" in Vienna 1943–45)

- Some archival objects refer to performance, e.g. photographs
- Connecting them to theadok.at because there it is possible to additional data



Handling of performance identifiers at tfm@Uni Vienna

Using theadok.at to:

- Collect metadata on performances that are relevant for the archive and collections at the tfm department using a dedicated data model for performances
- Refer to the theadok.at performance identifier in the different projects at the department that deal with performances, e.g. HIC and HoTS (= current project on the History of Theater Studies: Switzerland/Austria, funded by FWF and SNF)
- List the references to archive objects on theadok.at (but is not mandatory)

Integrating the performance identifiers in the software of the projects (in the input fields):

 Currently mostly copy & paste but also an autocomplete field is possible by using the API of theadok.at

No satisfying situation

What are the problems - local at tfm but also in general with current approaches:

- Complicated way to connect data and lot of steps to handle
- **Too much information** to handle
- Many different identifiers to combine: persons, buildings, companies, ...
- Distinguish between **performance plan and performance**
- Harvesting of data can be troublesome
- Handling of conflicting information is challenging
- Services should support FAIR data principles and high quality research data
- Complex data models and different interpretations of performances
- Vocabularies are not community-agreed and are under heavy discussion
- No accepted general authority file for performances/performance plans

Idea of a simple performance identifier system

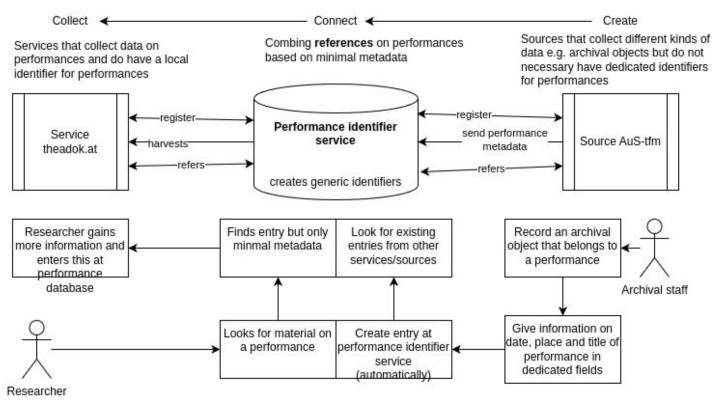
For the needs of tfm@Vienna we work on a prototype of a performance identifier system:

- Generate identifiers based on only few information: date title place
- **Services register** at this performance identifier system and integrate their data on performances: theadok.at and AuS-tfm as most important but also the mediathek of the department (collecting source and identifier at source)
- External identifiers are harvested, services register their data at the service
- Simple API focused on autocomplete functions to connect performances between the services
- Detailed information at the source not in the performance identifier system
- Support Linked Open Data approaches
- In short: the performance identifier system should act as a **mediator/faciliator**

Goals of a performance identifier system at tfm@Vienna

- **Lightweight approach**: use minimal data (title, place, date, source + identifier at source) and allow **ambiguity** (e.g. performance existed circa 1910)
- Create **simple identifiers** that point to an (once) existing entry
- Set a focus on harvesting of data from performance databases
- Allow changes and log them
- Keep a history of changes (if one entry was marked as typo but there are new ones coming with the same typo, it is maybe no typo - refer to the changes)
- Distribute from the system: not entering data in every service/database, only refer
 to the entry in the system with the help of the minimal data
- Decentralised services that can be connected (same concept and same approach)
- Upscale by adding more and more sources but also by connecting to the decentralised services => federated network

How could it work (sketch)



Challenges of the approach

- We don't use full information, e.g. no persons can made disambiguation tricky (looking into sources) but programming scripts could help
- Still, sources must be convinced to implement the minimal metadata input fields and to send changes to the identifier system
- It is now **all about the correct place**: title and date can be handled but place is complicated as it has a hierarchy inside (Burgtheater => Vienna => Austria) => need to define best practices in how to label a place
 - Language can become a problem
- Curation of data still necessary but should be easier, rule: change at source
 - o Flexibility is important: handle different kinds of granularity of data
- Still in discussion: How to **generate identifiers** in terms of **decentralised services** (the same combination of place-title-date should get the same identifier in every of the decentralised services)
 - Decentralised approach needs an common agreement on protocols and best practices

Takeaways

- Creating local identifiers is trivial if there is an entity "performance"
- **Challenge** is to **connect** these local identifiers and to handle the many cases where there is no dedicated entity performance available (as it is often the case at archives and collections)
- Unlikely that there will be an authority file for performances/performance plans, thus we need solutions to deal with this situation: decentralised approach seems to be helpful + harvesting of data + sharing and agreeing of protocols and best practices for minimal metadata
- Complex solutions can create more problems than anticipated, thus simple metadata and no fear for incomplete and incorrect information
- Curation and community work will stay and will be still important

Thanks for listening!

Contact: klaus.illmayer@univie.ac.at