The Dance Lesson by Edgar Degas (circa 1879), Public Domain

Using Wikidata for Performing Arts Related Data

Webinar of the OS-ADM Project, 5 June 2024

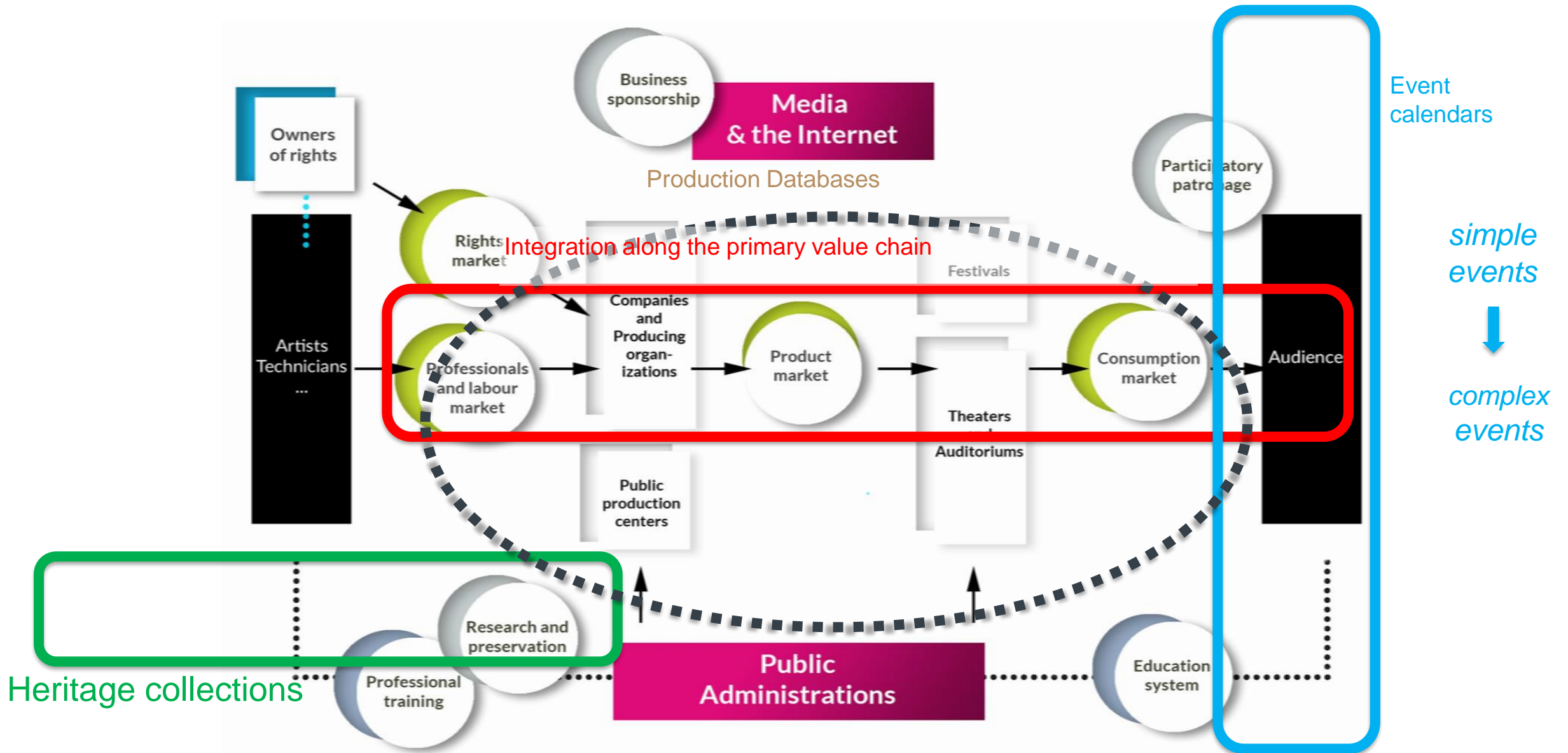Beat Estermann, Bern Academy of the Arts / Association Opendata.ch

Today's Programme

▸ **Part 1: Introduction and Overview**

- Overview and (possible) applications running on Wikidata
- Data model implemented on Wikidata
- Pros and cons of using Wikidata
- Recommendations regarding first steps

▸ **Part 2: Peer Consulting**

- Interactive Workshop

PA Value Network and Current Trends in Data Integration

Main Focus Areas of Today's Workshop

▶ How to interlink existing databases by using "bridge identifiers"?

▶ How to establish a collaboratively managed production database that can be used as a key finding aid to access event announcements, media coverage, and heritage collections?

▶ How to make performing arts related archives effectively searchable on the web through lightweight One-stop Shop solutions?

Data Interlinking and Integration via Wikidata

Main focus: Pragmatic data harmonization, interlinking, and integration across institutional boundaries

Documentation: WikiProjects "Performing Arts" and "Cultural Venues"

Tutorials: https://linkeddigitalfuture.ca/wikidata/

Applications

- Canadian Performing Arts Directory, entirely powered by open data (Artsdata.ca / Wikidata)

- Culture in Time App (facilitating exploration, combination and configuration of Linked Open Data sources)

- Swiss Archive of the Performing Arts (SAPA) working towards a "Panorama" Microsite entirely powered by Wikidata

Linking Datasets to Base Registers / Authority Files

Databases
- ISNI
- VIAF
- MusicBrainz
- Discogs
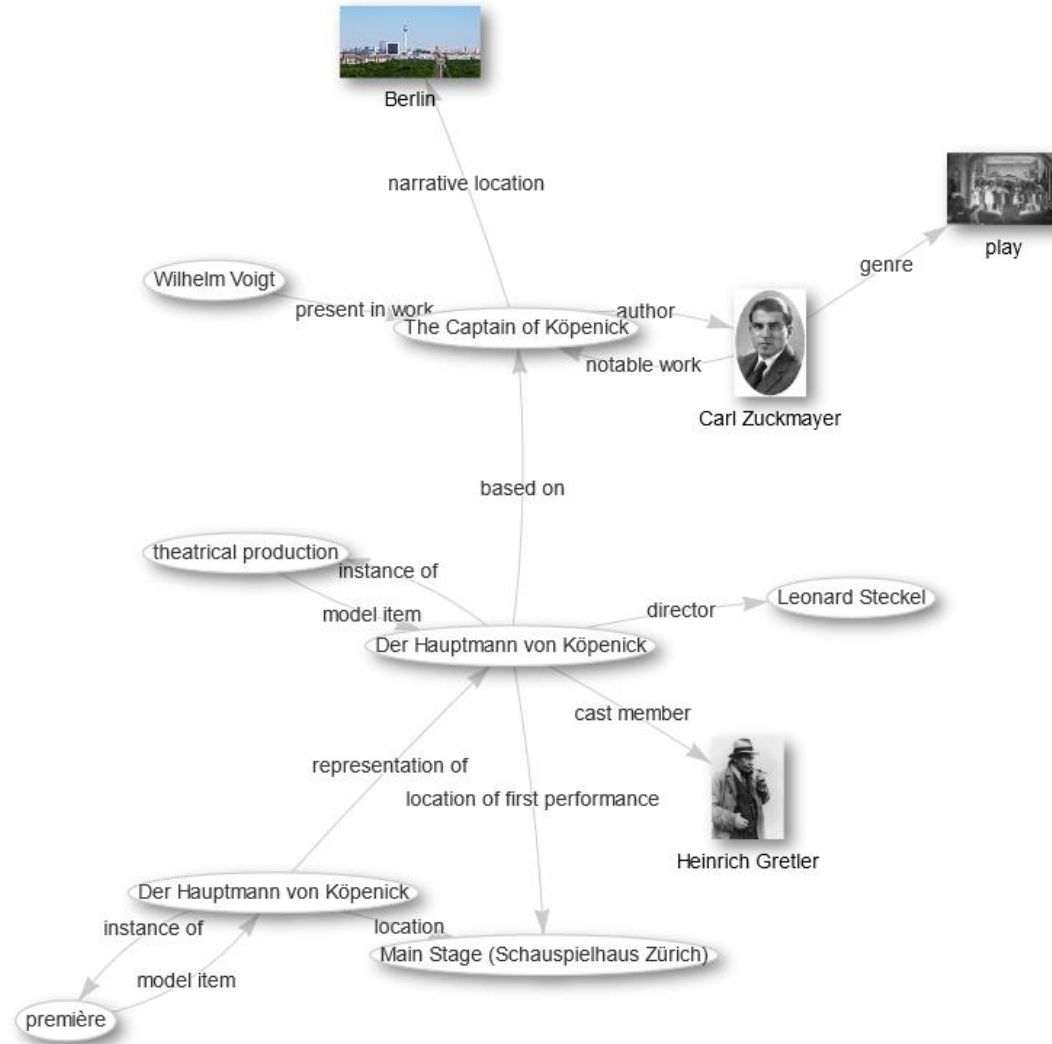- IMDb
- Songkick
- GND
- Wikidata

Entities
- Works *(literary, musical, choreographic)*
- Editions/Translations of Works
- Character Roles
- Performing Arts Buildings
- Humans *(writers, composers, performing arts professionals)*
- Organizations (*presenting organizations, musical ensembles, theatre ensembles, dance ensembles*)
- (Complex) Events (*performing arts productions, festivals, concert tours*)

Base Registers / Authority Files play a key role in **interlinking datasets** from various sources.

Some statistics (Wikidata, March 2024)
- 637'000 musical works
- 18'000 plays
- 1'200 choreographic works
- 21'000 character roles
- 16'500 performing arts buildings
- 380'000 musicians
- 420'000 actors/actresses
- 115'000 musical ensembles
- 14'000 theatre ensembles
- 1'500 dance ensembles
and steadily growing...

Ingesting Data About Performing Arts Productions

So far, data about approx. 26'000 productions have been ingested…

Example Items – Describing Performing Arts Productions

- **theatrical production**
  (in the original language, with labels for the character roles):
  Der Hauptmann von Köpenick (The Captain of Köpenick) (Q40289399)

- **theatrical production**
  (in a translated version, with labels for the character roles)
  Eine kleine Stadt (Our Town) (Q43689202)

- **guest performance (series of performances)**
  (in the original language, with links to the character roles):
  L'école des femmes (The School for Wives) (Q43759980)

- **premiere (single performance)**
  Der Hauptmann von Köpenick (The Captain of Köpenick) (Q39907209)

For further examples (e.g. for actors, character roles, works), see the project page.

Referencing Heritage Collections on Wikidata

Main focus: Light-weight referencing of heritage collections in Wikidata to facilitate cross-institutional search

Documentation: WikiProjects Heritage Institutions and Heritage Collections

Applications

- Wikipedia Infoboxes (French Wikipedia – overview; example)

- Graph-based query expansion on Library Discovery Systems (description), see pilot implementation: New Test View on Search Portal of ETH Library)

  Try searching for Angelique Rockas or Ettore Cella…

  Click on "Did you search for a person?" on top of the search results; select the respective person and click on "search for this person". At the bottom of the search results page, you will find a section "More information about this person"; this section includes references to archival holdings that have been entered on Wikidata as well as pointers to entries about the person in authority files.

Example Items – Referencing Heritage Holdings

(Super-)class of item	Type of artefact	Property used	Examples
human (Q5)	archival material	archives at (P485)	Angelique Rockas (Q23893984)
human (Q5)	documentation	documentation files at (P10527)	Bob Dylan (Q392)
human (Q5)	personal library	personal library at (P9419)	Tennessee Williams (Q134262)
organization (Q43229)	archival material	archives at (P485)	Flemish Theater Institute (Q1812913)
organization (Q43229)	documentation	documentation files at (P10527)	Stadttheater Zürich (Q110287453)

Source: https://www.wikidata.org/wiki/Wikidata:WikiProject_Heritage_Collections/Data_structure

Wikidata and Classical LOD are Complementary

	Wikidata	Classical LOD
Strengths	Fully-fledged crowdsourcing platform; further parties can easily be invited to contribute.	Data owners keep the control over their «graphs»; data quality and completeness remains under the control of the data provider.
	Immediate integration with the worldwide linked data cloud (reconciliation at the moment of data ingest)	Data can be published in RDF format, is linkable; reconcilation against other databases can be done step by step.
	Community-supported LOD service with a certain level of reliability	
Weaknesses	Monitoring data quality and completeness is a permanent and challenging task.	Third parties cannot readily fix issues related to data quality or completeness that are not taken care of by the data provider.
	Harmonization of data modelling practices is a challenge.	Harmonization of data modelling practices within one's own «silo» is straightforward, but might be a great challenge to implement across «silos».
	Perceived «loss of control» by data owners	Introducing collaborative data maintenance practices is difficult.
		Many current LOD services are of questionable reliability.

Prioritizing Your Data Publication Efforts

▶ When it comes to publishing data on Wikidata, **priority** should be given to data:

- where it is **unclear who would be the «natural» authority** in the given area (on a global scale);
- where there is a **high potential** for enhancing data through **crowdsourcing approaches** *(including community or expert sourcing)*;
- where there is a **high potential** for using the data to **provide linkages between many different datasets**;
- where data is likely to be reused in the context of **Wikipedia**;
- where **international coordination** to ensure semantic interoperability of the data is **unlikely to take place elsewhere**.

▶ It is best to focus on **base registers / authority files** and **controlled vocabularies** first; they facilitate further interlinking of datasets!

Peer Consulting Session

▶ Questions & Challenges of the Participants…

Example Case

▶ **Theater an der Winkelwiese (Zurich):**

- Enter data about current & upcoming productions in Wikidata to:

  - Provide a basic finding aid for related material uploaded to Wikimedia Commons
    (fully possible today)

  - Make the theatre's own databases fully interoperable with the production database of
    the Swiss Archive of the Performing Arts (SAPA)
    (partly possible today, depending on SAPA data ingests)

  - Link entries in third parties' event databases to the production database
    (about to enter the prototyping stage)

  - Allow journalists/publishers to link related media coverage to the production database
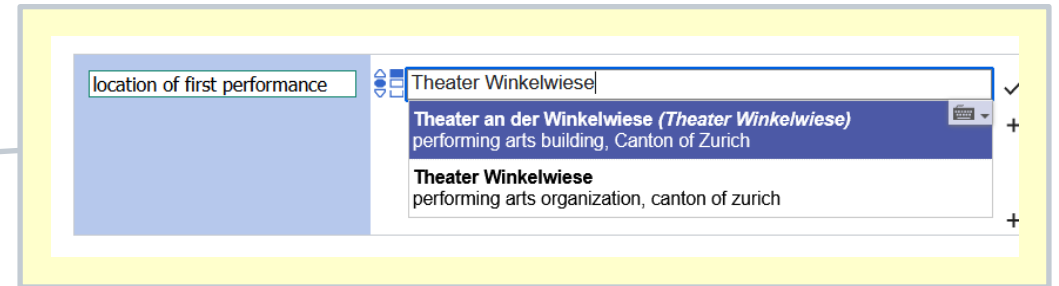   **(at the conceptual stage)**

Example Production: Sisyphos' Kinder (2024)

▶ **Source:** https://winkelwiese.ch/inszenierungen/sisyphos-kinder

▶ **Minimal Requirements for a production entry in a production database**
as discussed at the Theapolis Project "Inszenierungsdatenbank" (2022):

  ▶ **Basic identifying information:**
  Title, subtitle, first performance (date, place), season, production company

  ▶ Type of production (on stage vs. streaming; premiere type)

  ▶ Artistic contributions (contributors)

  ▶ Description, media upload

  Short description
  → Wikidata

  Long description, media upload
  → Wikimedia Commons

Data Ingest – Step by Step ([Wikidata Item](#))

- ▶ type (instance of)
  → "performing arts production" *(or one of its subclasses, e.g. theatrical production)*
- ▶ title
- ▶ subtitle
- ▶ date of first performance
- ▶ location of first performance → **Venue**
- ▶ production company → **Organization**

- ▶ genre → [Controlled Vocabulary](#)
- ▶ has premiere type → [Controlled Vocabulary](#)

- ▶ director
- ▶ scenographer   → **Person**
- ▶ cast member
- ▶ …

- ▶ based on → **(Literary) Work**

Contact

Beat Estermann

Opendata.ch
Member of the Board

beat.estermann@opendata.ch
https://openglam.ch

Bern Academy of the Arts
Coordinator Open Science / Digital Humantities

beat.estermann@hkb.bfh.ch

On LinkedIn