





# Datengrundlage

Theapolis

- ~ 24.000 Künstlerische Leitungen
- ~ 20.000 Von (Autor\*innen, Urheber\*innen, ...)
- ~ 18.000 Organisationen
- ~ 90.000 Produktionen
  - davon ~ 7.000 Produktionen der freien Szene



# Datenaufbereitung

Wikidata + GND

- Matching der Personen
- Matching der Organisationen
- Matching der Produktionstitel mit Werken unter Berücksichtigung der Autor\*innen/Komponist\*innen/Regisseur\*innen/...
- Reconciling in mehreren Iterationen gegen Wikidata und GND
- **Ziel:** Zuordnung der Wikidata Q Identifier und/oder GND ID



# Datenmodell

Schema.org

- Basierend auf Schema.org
  - **CreativeWork** (für Produktion + Werk)
  - **Person**
  - **Organization**



# Datenmodell

Wikidata

- Erweiterung der Produktionen um relevante Properties aus den entsprechenden Wikidata Entitäten
  - performing arts production
  - dance production
  - theatrical production
  - ...



# Datenmodell

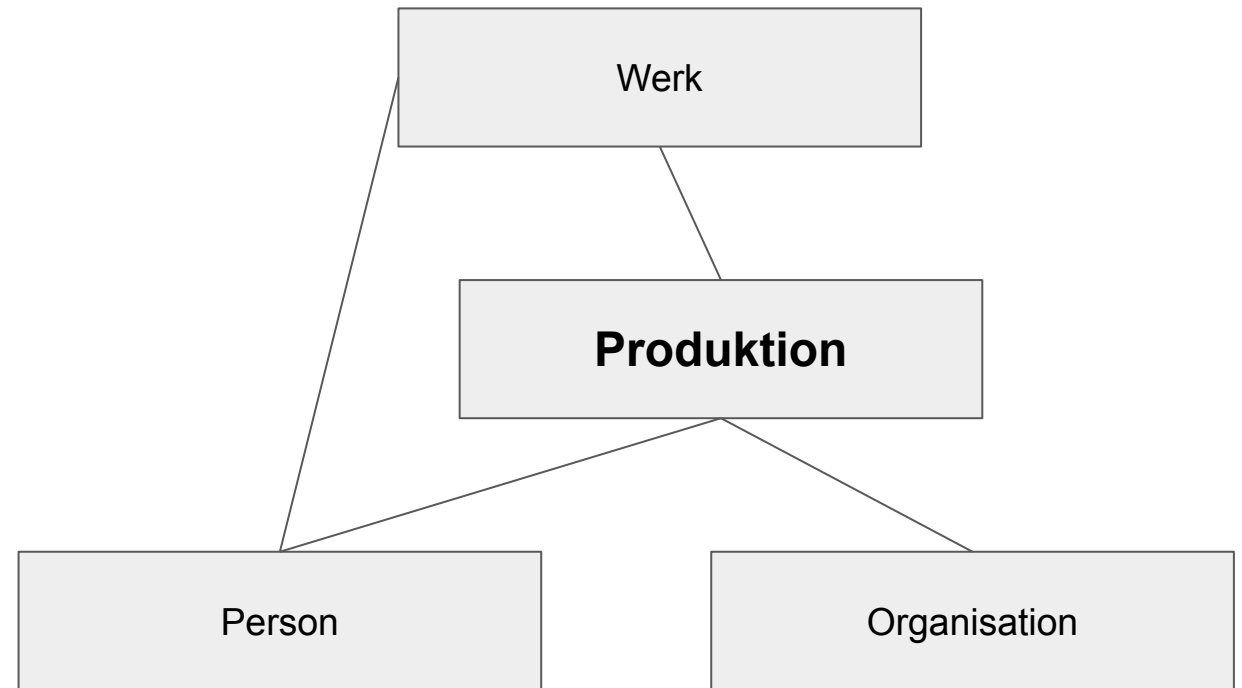
Spezifische Anforderungen

- Erweiterung um spezifische Felder, für die Workflows innerhalb der Produktionsdatenbank
  - Status der Produktion
  - Duplikatsverweise
  - Externe Identifier
  - ...



# Datenmodell

Skizze





# Datenmodell

## Identifizier

- Eindeutige Identifizier
  - Wikidata Q Identifizier
  - GND ID
- Externe Identifizier
  - Theapolis Identifizier
  - (Berlin Bühnen Identifizier)
- Zuordnung mehrerer externer Identifizier
- Persistente Identifizier für alle Entitäten
  - Keine Löschung sonder immer ein Verweis auf die neu zugeordnete Identität





# Datenbank

Drupal

- Umsetzung des Datenmodells in Drupal (Open Source CMS)
- 1. Iteration
  - Eingabeformulare
  - Detailansichten zu Produktionen, Werken, Personen und Organisationen
  - Listenansicht der Produktionen mit Suche und Filtern



# Datenbank

JSON:API

- 1. Iteration
  - JSON:API für Produktionen, Werke, Personen, Organisationen
  - Zugriff auf sämtliche Inhalte mit Such- und Filterfunktionen
  - CRUD Endpoints



# Migration

## Theapolis

- Migration der angereicherten Theapolis Daten
- Nutzung der JSON:API
  - Lookup Prozesse zur Vermeidung von Duplikationen (Fallback QID, GND, Externer Identifier, Name)
- Zusätzliche API als Middleware für den kontinuierlichen Import von Daten aus Drittsystemen



# Migration

Stand Heute

- ~ 35.000 Personen
- ~ 17.500 Organisationen
- ~ 3.500 Werke
- ~ 7.000 Produktionen der freien Szene



# Kontinuierliche Migration

Theapolis + Berlin Bühnen + ...

- Mapping über externe Identifier
- Abgleich von Neueintragungen mit Wikidata + GND
  - 100% Treffer > Direktübernahme
  - weniger als 100% Treffer > Flag für Redaktion



# Aufbereitung Datenbestand

Wikidata + GND + ...

- kontinuierliches Erweiterung um fehlende Information durch Wikidata + GND Abgleich
- kontinuierliche Plausibilitätsprüfung und Flagging für manuelle Nachbearbeitung
- Direkte Integration beziehungsweise Verlinkung von Partner-Inhalten



# User Interface

TODO

- vereinfachung der vom CMS vorgegeben Formulare als Assistent zur Eintragung von Produktionen
- Zuweisung weiterer Mitarbeiter\*innen für die Produktionen
- Dialog durch kontinuierliche Benachrichtung