

Protokoll 1. Beteiligungsforum Inszenierungsdatenbank Technik-Gruppe

Meeting am Freitag, den 24.06.2022

Anwesende:

Janine Thoenelt und Patrick Krüger (Theapolis), Prof. Beat Estermann (Berner Fachhochschule), Julian Kamphausen (Studio für unendliche Möglichkeiten), Livia Rutishauser (DTD), Patrick Primavesi (Universität Leipzig), Rainer Glaap (ex-Theaterportal), Rüdiger Koch (Papiertheater Invisius)

Anne Müller (Protokoll)

Glossar:

Inszenierungsdatenbank (InszDB)

Vorstellungsrunde

Janine Thoenelt (Theapolis)

- Geschäftsführerin Theapolis und Moderatorin der Technikgruppe

Julian Kamphausen (Gründer Studio für unendliche Möglichkeiten)

- Anlass zur Teilnahme: Konzeption und Umsetzung eines (nichtlinearen) Archivs für Theaterkollektiv Pilkentafel in Flensburg
- Erfahrungen in Front- und Backend Entwicklung, mit dem Ziel, die Nutzung und Bedienung dieser Tools zu vereinfachen → Tools und Software kann von Projekten selbst bedient werden (möglichst OpenSource, Datenschutzkonform)

Beat Estermann (Berner Fachhochschule)

- an der Berner Fachhochschule
- Institut Public Sector Transformation
- Themen: Gedächtnisinstitutionen, digitale Transformation, OpenData, Hackathons
- seit 2013 Zusammenarbeit mit der Schweizerischen Theatersammlung
- Kontakt mit nationalen und internationalen Playern bzgl. Performing Arts Daten

Patrick Primavesi (Universität Leipzig)

- Theater- und Tanzwissenschaftler an der Universität Leipzig
- Verantwortlich für das Tanzarchiv Leipzig (wird mittlerweile an Universitätsbibliothek verwaltet) → Thema: Zugänglichkeit der Datenmengen für Recherche
- Großes Interesse an Vermittlung zwischen Wissenschaft und Praxis
- Zusammenarbeit mit anderen Tanz-Archiven in Deutschland
- Wie verbessert man Datenmodelle, wie kann damit gearbeitet werden und diese sichtbar machen → "Silodenken" hilft nicht weiter
- Künstler*innen, Produktionsstätten, Wissenschaft, Veranstalter und Ensembles sollen sich angesprochen fühlen

Patrick Krüger (Theapolis)

- Seit 24 Jahren Webentwickler, UX Designer und UI
- Seit anderthalb Jahren Projektmanager bei Theapolis → Leitet das Technikteam und betreut technische Weiterentwicklung

Livia Rutishauser (DTD)

- Koordination der Plattformentwicklung beim Dachverband Tanz digital → Plattform für Tanz im Digitalen Raum
- Aktuelles, zeitgenössisches und Archivmaterial sollen in Screening-Angeboten miteinander verknüpft werden
- → Hoffte, ein neues Verständnis für die technischen Grundvoraussetzungen zu erhalten und sieht die größten Herausforderungen in den strukturellen Daten

Rainer Glaap (*ex-Theaterportal*)

- gelernter Theaterwissenschaftler
- 36 Jahre in IT Industrie tätig gewesen
- 2002 Gründung Theaterportal → Ziel: Übersicht, Buchungsplattform, Contentserver für Rezensionen, Lebensläufe etc. → eine Art IMDb der Bühnen/Darsteller*innen
- 15 Jahre bei Eventim → Betreuung von Opern- und Theatersoftware und immer enger Kontakt zu den Institutionen

Rüdiger Koch (Papiertheater Invisius)

- Puppenspiel, Studium Theatertechnik
- Für Schaubude und Kulturprojekte Berlin gearbeitet
- 20 Jahre in der Veranstaltungswirtschaft tätig
- Blog: <https://publikumsschwund.wordpress.com>
- → Daten sollten bei Wikidata abliegen, Normdaten sind enorm wichtig
- → Möchte auf ein weiteres Pilotprojekt hinweisen: Stadtmuseum Berlin hat 17.000 Datensätze freigeben und möchte sich beteiligen (AP: Bärbel Reißmann)

Basisinformationen

Theapolis stellt im Chat Umfrageergebnisse zu folgenden Fragestellungen zur Verfügung:

- Wie werden die Daten gespeichert?
- Woher werden die Daten erhalten? Wer trägt ein?
- Wie wird die Datenqualität verifiziert?
- Szenarien und Erwartungen

Themensammlung und Diskussion

Patrick Krüger:

- Grundsatzfrage: Können wir sofort mit Beispieldatensätzen arbeiten oder wären folgende Fragen im ersten Schritt wichtig zu klären:
 - Tragen alle Beteiligten alle Daten direkt in WikiData ein oder ist die InszDB eine Zwischenebene, die die definierten Kerndaten erfasst und diese dann

wiederum bereinigt und mit bestimmten Qualitätskriterien dann an WikiData schickt.

- Gruppe muss sich, noch vor der konkreten Beschäftigung mit Wikidata, mit allgemeinen Entscheidungen befassen.

Patrick Primavesi

- Was ist aus Sicht der Interessengruppen machbar und wie könnte die Pflege der InszDB mit bisherigen Routinen verknüpft werden, ohne neue Arbeitsfelder aufmachen zu müssen und ohne Mitwirkende abzuschrecken?
- Es sollten in dieser Runde verschiedene Stakeholder Perspektiven versucht werden mitzudenken.
- Es muss bei Interessensgruppen ein Bewusstsein für die vorhandenen Daten geschaffen werden, damit auch Datenbestände nicht verloren gehen/gelöscht werden
- Ziel: Es sollte sich auf ein Datenmodell geeinigt werden, wo auf elementarer Ebene alle zusammenkommen (trotz unterschiedlicher Interessens- oder Anwendungsfelder)

Janine Thoenelt

- zu den Vorrednern: Für große Datenplattformen sollte es Schnittstellen geben, um die Daten einzuspeisen, sonst droht Abschreckung.

Rüdiger Koch

- Erfahrungen aus dem letzten Jahr zeigen: Wikidata ist für viele nicht gut zugänglich und nicht verständlich → Es bedürfte Schulungen
- Dateneingabe ist auch zu viel für Institutionen
- ist es Möglich eine Schnittstelle zu Wikidata zu haben und darüber hinaus Daten auch selbst einzugeben?

Rainer Glaap

- Woher kommen die Daten für eine gemeinsame InszDB? Datenquellen könnten z.B. auch Ticketsysteme sein, die unmenge an Daten besitzen (z.B: EVENTIM.Inhouse hat Daten bis 1996). Dort gibt es Schnittstellen und Möglichkeiten Basis-Daten zu exportieren und nutzen: Bühnen besitzen die Datenhoheit und können diese an andere Plattformen ausspielen
- In England gibt es hierfür bereits zwei große Initiativen, wo via Schnittstellen täglich Daten zentral gesammelt werden

Beat Estermann

- stimmt Vorrednern zu:
 - Wikidata ist vermutlich nicht der geeignete Ort zur Datenmodellierung/-sammlung aller Akteur*innen
 - Prospektiver Ansatz: Wir bauen Erfassungsmasken, die in Wikidata inkludiert werden.
- Neben Schnittstelle kann auch eine individuelle Datenbearbeitung möglich gemacht werden / muss implementiert werden (z.B. Intendanzwechsel) → für solche Anwendungen müssen Tutorials entstehen oder Workshops angeboten werden

Patrick Krüger

- Bei großen bestehenden Datenbanken könnte der return on invest hoch sein wenn eine automatische Konvertierung bzw. ein Abgleich der Daten geschaffen wird. (XML Files via Schnittstelle zu Wikidata)
- Was ist mit den anderen Playern, die strukturierte Daten gar nicht nutzen/erzeugen (z.B. unstrukturierte Wordpress-Seiten) → Wie gehen wir hiermit um? Was soll das Projekt hier leisten?
- Große Fragen: Wie kommen die Daten in die InszDB? Welche Tools müssen wir den Nutzer*innen an die Hand geben, damit InszDB nutzbar wird und Nutzung Spaß macht?

Rainer Glaap im Chat

“Historische Daten aus einem Ticketingsystem: Seit den 80ern gab es System namens MUETHOS, von einem Herrn Böttcher erfunden und betrieben. Es gibt nur noch wenige Installationen, möglicherweise könnte man mit B. darüber reden, dass er Daten exportiert ... Ausserdem beendet gerade die Bundestheaterholding in Wien die Nutzung von CULTURALL. Der Betreiber ist an der Umstellung beteiligt. Ev. kann man über ihn auch einen Datenexport der letzten xx Jahre bekommen ...”

Livia Rutishauser

- Frage: Müssen immer erst Daten händisch eingetragen werden, um dann damit die API entwickeln/festlegen zu können?
- Status quo Tanz digital: welche Daten sind übereinstimmend notwendig → im nächsten Schritt wird festgelegt, welche Daten müssen importiert oder exportiert werden

Beat Estermann

- Erster Schritt: Abgleich welche Daten sind vorhanden und Abgleich Daten Wikidata
- Strukturierte Datensätze müssen erst bereinigt werden: Patrick Krüger und er haben bereits einen Test gemacht und haben festgestellt, dass Datenbanken beispielsweise Spielstätten und Orte verwechseln
- Matchen der Einträge ist wichtig, um keine Dubletten zu erzeugen → Abgleich mit Wikidata muss über Tools möglich sein
 - Tool für kleine Datensätze: Open Refine → Reconciliation Abgleich
 - Jedoch wird das Match statistisch ausgewertet und bei bspw. einem 50% Match, muss der Mensch entscheiden/redaktionell eingreifen

Rüdiger Koch

- Open Refine ist für allgemeine Nutzung nicht geeignet.
- Feststellung: Daten in Wikidata sind, bis auf die von Beat Estermann, nicht sehr valide.
- Es braucht einen Anreiz für die Gedächtnisinstitutionen, die Daten aufzubereiten

Patrick Krüger:

- Es muss eine einmalige Übertragung der Daten an Wikidata geben und auch das bereinigen über Open Refine muss stattfinden.
 - Aber wie wird mit weiteren Aktualisierungen verfahren?

- Wikidata vergibt genormte, eindeutige QIDs → mit diesen Daten muss entsprechend weiter gearbeitet werden können.
- Die InszDB könnte der Mittler sein, das Register für Basisdaten.
- Es braucht diesen definierten Basisdatensatz. Darüber hinaus können Institutionen dann selbst entscheiden Daten zu verknüpfen und hinzuzufügen. So kann auf unterschiedliche Ansprüche und Befindlichkeiten eingegangen werden.

Patrick Primavesi

- Mittelfristiges Ziel: Bewusstsein und Vernetzung schaffen → Paralleluniversen abschaffen bzw. bündeln
- Es gibt Initiativen die ähnliche Ansätze nutzen wie soeben von Patrick Krüger beschrieben:
 - Projekt Gemeinsame Normdatei (GND) Bibliotheken Datenbank:
 - Arbeitsgruppe arbeitet an Unterscheidung zwischen Basisdaten (Werknormdaten) einer Publikation und was sind Zusatzinformationen durch unterschiedliche Quellen (und wie werden diese in der Datenbank angesiedelt und abrufbar gemacht)
 - Ziel: Zweifelsfreie Unterscheidung zwischen Werken
 - Datum, Beteiligte, Körperschaft scheinen die wichtigsten Basisdaten zu sein
- Referenzierbare Basisdaten müssen möglich sein. Nur so entsteht ein Mehrwert für die Community.
- späterer Schritt: Auch andere vorhandene Daten sollten verknüpft werden können und zugänglicher werden.

Beat Estermann

- Stimmt Patrick Primavesi zu: Ein Minimaldatensatz für Normdaten muss definiert werden und ist notwendig

Patrick Krüger

- Sind z.B. Spielzeiten und Vorstellungstermine relevant?

Rüdiger Koch

- Kleinster gemeinsamer Nenner sollte gefunden werden. Die Vorstellungsdaten sollten daher eine nachgeordnete Rolle spielen
- Erst mal Fokus auf Produktionen

Patrick Primavesi

- Premierendatum ist wichtigster Parameter, sonst sind Produktionen nicht korrekt auffindbar und referenzierbar
- Außerdem: Körperschaften und Beteiligte

Beat Estermann

- Saison, Regisseur und Titel würden ausreichen

Patrick Primavesi

- Saison ist nicht ausreichend, diese gibt es nicht in der Freien Szene → daher Entscheidung für Premierendatum
- Vorstellung / Premiere sollte mit dem Datensatz der Produktion verknüpft sein → somit kleinster gemeinsamer Nenner und wenigstens eine Veranstaltung ist im System
- Datenbank Pina Bausch ist ein sehr gutes Beispiel für fortgeschrittene Datensammlung → aber das ist nur Anregung für Zukunft
- Ziel: Identifizierbarkeit der Produktionen

Rüdiger Koch im Chat

“factgrid ist eine Wikibase-Instanz, die sich anbot ebenfalls zu hosten... oder wir machen eine eigene Wikibase-Instanz”

Julian Kamphausen

- Um Schwelle für ehrenamtliche Mehrarbeit zu senken, sollten Variablen eingegeben werden können, z.B. Premierendatum oder auch nur das Jahr
- es gibt zwei Mögliche Wege zur Pflege der Datenbank: 1) Mit Beteiligung der Institutionen/Künstler*innen und 2) ohne deren Beteiligung
- kennt die Szene: Daten aus amerikanischen Unternehmen wie Google und Eventim (auch Wiki Foundation) schrecken die Player ab. Das sollte bedacht werden.

Rüdiger Koch

- Wo hosten wir die Daten?
- FactData kommt nicht in Frage → bleibt eine potenziell sinnvolle Zusammenarbeit via Wikidata

Janine Thoenelt

- Erfahrungswerte bei Theapolis zeigen, dass je weniger Details abgefragt werden, um so höher die Chance, dass Daten eingetragen werden.
- zum Thema Vernetzung schaffen: Kann sich vorstellen, dass die GND beispielsweise in Zukunft mal die Instanz sein könnte, die später eine Verifizierung der einzelnen Produktionen vornimmt und eine “eigene” ID vergibt

Zwischenfazit der Arbeitsgruppe

Beat Estermann

- Es bietet sich an, mit bestehenden Daten einen Testballon zu starten
- Parallel zu diesem Prozess ist für die prospektive Eingabe der Daten zu klären:
 - Wo können Daten eingegeben werden?
 - Zentrale oder dezentrale Systeme?
 - Wo und wie werden die Daten ausgespielt?
- ohne diese Requirements, kann Programmierung nicht beginnen

Julian Kamphausen

- Prognose: Schnittstellen zu Google oder Eventim wird in der Freien Szene auf Ablehnung stoßen, Diskussion zu diesem Thema muss zeitnah geführt werden

- Es sind sehr hohe Projektkosten zu erwarten, wenn es keine Incentives für die Protagonist*innen gibt, sich selbst daran zu beteiligen
- Akzeptanz in der Szene wird geschwächt

Rüdiger Koch

- Google als Player ist nicht zu verhindern, denn sobald Daten "greifbar" (Open Data) werden eigene Datensätze generiert

Patrick Primavesi

- Die Vorbehalte, die es in der Runde gibt, sind ernst zu nehmen
- Es sollten weitere Kommunikationsschritte eingeplant werden → Klärung der Fragen, Worum handelt es sich? Sinn und Mehrwert des Projektes? Risiken und Mehrwert (Kommerzialisierung und Datenschutz)?
- Das Wissen und die Erfahrungen um Datenkomplexität, ist sehr wichtig, um gleiche Fehler nicht nochmal zu machen
- Julian Kamphausens Bedenken sind ernst zu nehmen: es muss genau kommuniziert werden, was sind die Zielstellungen, Chancen und für welche Bereiche, ist die InszDB sinnvoll → nur so können Missverständnisse vermieden werden

Janine Thoenelt

- Sollte ein Thema für das kommende Beteiligungsforum sein: Wie bringt man Player zur Nutzung der InszDB (Kommunikation, Marketing/Werbung)

Patrick Krüger

- Ohne Kritische Masse machen neue Netzwerke keinen Spaß
- Bottom-down-Ansatz: vorgegebene Strukturierung der Dateneingabe an die User*innen
- Zusammenarbeit mit den Landesverbänden sehr wichtig
- Identifikation: Was sind Minimaldaten
- Welche Daten dürfen/können überhaupt veröffentlicht werden → Schützenswerte Daten
- Wer wird Daten liefern? User Generated Content? Community Projekt?

Livia Rutishauser

- Stimmt Patrick Krüger zu → Ein Netzwerk ist dann relevant, wenn es groß genug ist, um lebendig zu sein und einen Mehrwert bildet
- Fragen bleiben: Datenspeicherung, Server, Datenschutz (Vorbehalte gegenüber US-amerikanischen Unternehmen)
- Welche Initiative steckt dahinter, z.B. europäische, deutschlandweite → auf Mentalitäten muss reagiert werden können

Beat Estermann

- Datenfreigabe: Es kann mit Opt Out Funktionen gearbeitet werden

- Wikidata hat eigene Policy für Datenfreigabe, so kann Datenveröffentlichung ebenfalls gesteuert werden.

Patrick Krüger

- Zusammenfassend:
 - Definition der Kerndaten muss abgewartet werden (z.B. Produktion, Company, UA)
 - Diese Daten sind wahrscheinlich zu großen Teilen vorhanden und teils öffentlich
 - Es ist den Beteiligten freizustellen, welche weiteren Informationen, durch wen geliefert werden und wie diese außerhalb der InszDB zur Verfügung gestellt werden (öffentlich, Paywall etc.)
 - Über wenige Entitäten diskutieren und daraus Minimaldaten identifizieren
 - ID vergabe via Wikidata muss erfolgen, um Register aus Basisdatensätzen erstellen zu können (QID Vergabe)
 - allein dieser Vorgang ist ein umfangreiches, aufwendiges Projekt

Zusammenfassung

- Arbeitsgruppe auf die Zuarbeit und Ergebnisse anderer Arbeitsgruppen angewiesen
- Priorität: Einigung auf Kerndaten (auch Basisdaten, Minimaldaten) sowie Anwendungsszenarien
- Es braucht einen definierten Basisdatensatz und darüber hinaus können Institutionen selbst entscheiden Daten zu verknüpfen und hinzuzufügen. So kann auf unterschiedliche Ansprüche und Befindlichkeiten eingegangen werden.
- Set von Kerndaten sollte der Fokus sein, aber nicht zu klein denken, denn sonst gibt es keinen Nutzen dafür → zu allgemeine Daten reichen auch nicht aus
- Dateneingabemasken vergleichen um Aufwand der Zusammenführung der Daten identifizieren zu können. → In der Theapolis-Umfrage wurden bereits die Felder angefragt. Es zeigt sich schon, dass die Daten sehr unterschiedlich sind.
- Fazit: Es sollte am praktischen Beispiel Datenlage geprüft werden
- Es gibt bereits ein angesetztes Treffen Hands-on Wikidata-Workshop am 6.7.2022 (13:00 - 15:30)
- Konkrete Ingest-Fragen können dort geklärt werden indem Datensätze exemplarisch eingespeist und verglichen werden
 - Organisiert von FID Darstellende Künste in Kooperation mit der Berner Fachhochschule
 - Konkreter Übungs-Fall: Produktionsdatenbank Figurentheater (Rüdiger Koch)
 - Weitere Fragen aus konkreten Pilot-Ingests können ebenfalls beantwortet werden.

- → Interessierte sind herzlich eingeladen (Anmeldung: https://docs.google.com/document/d/1IXYKEn8u3i0RYyx_UEBI9zRqgscXYjv4Hr552e-3lXA/edit#)
- Umfrageteilnehmer*innen, die angegeben haben, Daten zu besitzen, könnten gezielt angesprochen werden.
- Wikidata
 - Es muss eine gesonderte Daten-Eingabemaske erstellt werden, denn Wikidata ist für normale User*innen nicht geeignet. → Tool, welches Daten aufbereitet übergibt
 - Wikidata vergibt genormte, eindeutige QIDs → mit diesen Daten muss entsprechend gearbeitet werden um Register erstellen zu können
 - Liegt diese Daten-Eingabemaske zentral oder dezentral bei den Akteur*innen?
- Datenschutz ist ernst zu nehmendes Thema und muss diskutiert werden
- Ein Netzwerk ist dann relevant, wenn es groß genug ist, um lebendig zu sein und einen Mehrwert bildet → Kritische Masse bedenken
- Es braucht Anreize für die Gedächtnisinstitutionen und Akteur*innen, die Daten aufzubereiten
- Zeitplan des Projektes muss noch festgelegt werden

Dokumente

- [Impulsreferat](#) (Patrick Krüger)
- [Impulsreferat](#) (Prof. Beat Estermann)
- [Wikidata - Wikiproject Performing Arts](#)
- https://docs.google.com/document/d/1IXYKEn8u3i0RYyx_UEBI9zRqgscXYjv4Hr552e-3lXA/edit#