

Sluttrapport Arkivverket Arkivspira

Ref: 2023/14327

Prosjekttittel

Linket data og scenekunstarkiv. Automatisk innsamling av metadata fra 7 repertoardatabaser.

Tilskuddsmottaker

Stiftelsen PAHN – Senter for Scenekunst Norge (tidligere: Danse- og teatersentrum)

Ingress/oppsummering av prosjektet

Ved hjelp av midler fra Arkivverket har vi i 2024 gjennomført en pilot hvor vi har linket 6 databaser til Scenewebdatabasen; 1 kalenderdatabase (scenekunst.no), og 5 institusjoners turné- og repertoarverktøy (Riksteatret, Teater Innlandet, Teater Ibsen, Brageteatret og Turnéteatret i Trøndelag). Piloten har gitt oss god kunnskap i metodikk knyttet til automatisk innsamling av data. Vi vil ta med oss erfaringene fra piloten når vi nå videreutvikler denne til ett 3-årig prosjekt.

Startdato

4. januar 2024

Sluttdato

31. desember 2024

Generelt om prosjektet:

Repertoarhistorikk er en sentral del av scenekunstens historie. Den ligger som et grunnfjell i teaterhistorisk forskning. Det er kun ett sted brukere kan finne en nasjonal oversikt over scenekunstrepertoar i Norge: Sceneweb. Det er derfor viktig at vi dokumenterer repertoarhistorikken grundig slik at Scenewebdatabasen kan være en nyttig ressurs for forskere, studenter, markeds- og informasjonsmedarbeidere ved teatrene, journalister o.a.

Repertoarhistorikken fungerer på mange måter som Scenewebdatabasens skjelett, med digitale filer (fotografier, plakater, programmer, video, artikler, anmeldelser, magasiner o.a. digitalt dokumentasjonsmateriale.) bygget rundt repertoaret. Store ressurser går med til å bygge dette skjelettet i Sceneweb.

Ved hjelp av midler fra Arkivverket gjennomførte vi en pilot i 2024 hvor vi linket 6 databaser til Scenewebdatabasen; 1 kalenderdatabase (scenekunst.no), og 5 institusjoners turné- og repertoarverktøy (Riksteatret, Teater Innlandet, Teater Ibsen, Brageteatret og Turnéteatret i Trøndelag). Slik ønsket vi å teste ut metodikk for automatisk innsamling av repertoardata.

Mål

Hovedmålet var å teste ut metoder for automatisk innsamling av data fra digitale arkiv produsert av kulturbransjens egne aktører. Ved hjelp av linket data metodikk ønsket vi å koble 7 eksisterende databaser til Scenewebdatabasen. Disse databasene inneholder informasjon om teater- og danseoppsetninger, hvilke personer som deltar i disse, og hvor og når oppsetningene spilles, noen inneholder også fotografier, plakater og videoteasere.

Målet var å automatisere og effektivisere innsamling av metadata, unngå dobbeltarbeid, og et mer oppdatert arkiv til en rimeligere kostnad.

Metodikk

Siden dette var ett prosjekt som skulle vare kun ett år og hadde et begrenset budsjett, gjorde vi noen valg i den opprinnelige prosjektbeskrivelsen som vi vurderte som viktige for å kunne gjennomføre koblingene, slik at vi var sikre på at vi fikk testet metodikken grundig. Vi valgte teatre som kjenner godt til oss og våre IT-utviklere, da teatrene benytter samme system som også driftes av våre IT-utviklere. Scenekunst.no er også en aktør som kjenner oss godt og vi har kommunisert mye med. Vi brukte derfor lite tid på å skape tillit mellom partene. Dette var essensielt for å få erfaring med flere aktører/teatre og databaser på så kort tid til en så rimelig pris (se vedlagt regnskap).

Prosessene startet med dialog med ledelsen hos eierne av databasene; teatrene og scenekunst.no. Vi forklarte hva prosjektet gikk ut på, hva og hvorfor vi ønsket å gjøre dette arbeidet, og hvilke fordeler det ville ha for dem. Det var også viktig for dem at dette ikke medførte noen utgifter for dem. I innledende møte snakket vi også om hvordan de brukte sine databaser og hva slags data som lå i disse. Vi diskuterte personvern, rettigheter og viktigheten av at eierne var bevisste hva slags data de kunne dele fra sine databaser. Da teatrene benytter disse databasene som interne planverktøy ligger det også sensitivt materiale i disse.

Neste steg var å inngå en kontrakt. Kontraktene beskriver i detalj hvilke data som skal eksporteres til Sceneweb, eierskap og opphavsrett til filer og metadata, tid for eksport av data, rett til tilgjengeliggjøring av dataene på Sceneweb.no (se vedlagt eksempel på kontrakt).

Når kontrakten var underskrevet av begge parter ble ett API utviklet av IT-utvikleren til eieren av databasen som skal eksportere data til Sceneweb. Scenewebs IT-utvikler linket deretter databasene sammen ved hjelp av dette API'et. Hver import gikk gjennom to runder med testing før de ble koblet til.

I Scenewebs backend har hver import sin egen plass. Der kommer produksjonene til syne når de er klare for import. P.g.a. mulige forandringer i personal og spilleliste blir dataene eksportert når produksjonene er ferdigspilt, for slik å få riktigst mulig data. Vi må manuelt godkjenne hver produksjon som skal importeres. Dette er for å kvalitetssikring av data, slik at vi unngår blant annet skrivefeil og opprettelse av duplikater.

Koblingen av data (spillesteder, personer og produksjoner) fra hver database huskes automatisk. For eksempel ved en produksjon som tas opp igjen: databasen vil da «huske» at dette er samme produksjon og kun legge til forestillinger og spillesteder på spillelisten, og

eventuelt nye medvirkende personer i den nye spilleperioden. Den «husker» også personer og spillesteder etter at disse er koblet/godkjent manuelt en gang. F.eks. vil personen Petronella Barker måtte godkjennes kun en gang for hver database - altså 5 ganger (om hun spiller på alle disse 5 teatrene).

Annet arbeid som gjøres manuelt: Vi legger også til noe mer metadata manuelt, for å matche vår registreringsstandard. F.eks. metadata på filer (foto, plakater, video), sjanger og tags, koble til originalverk.

Innsamlet data

Vi får inn de dataene som aktørene selv legger inn i sine systemer. Det er variasjoner i hvordan aktørene bruker sine databaser. Databasene inneholder informasjon om teater- og danseoppsetninger, hvem som er hovedprodusent(e) bak produksjonen, hvilke personer som deltar i disse, og hvor og når oppsetningene spilles, en av databasene inneholder også fotografier, plakater og videoteasere (Teater Innlandet), en av databasene inneholder ett fotografi per produksjon (scenekunst.no).

Evaluerings

Piloten har gitt oss gode erfaringer og kunnskap i metodikk knyttet til automatisk innsamling av data. Dette vil vi bringe inn i videre arbeid:

- Det er sentralt å bruke god tid og være tålmodig i innledende samtaler med eiere av databaser. Flere aktører hadde behov for god tid til å forstå hva deling av data innebærer, og for å tenke seg om og komme frem til hvilke data de ønsket å dele. Kulturaktørenes databaser er interne planverktøy som også inneholder sensitiv informasjon, hvor personvernloven er gjeldende, og det er viktig at de er trygge på sikkerheten rundt koblingen til Sceneweb slik at dette ikke deles.
- Når dette fundamentet var på plass, opplevde vi både stor velvillighet for deling av data og forståelse av viktigheten med oppdaterte arkiv i Sceneweb.
- Det var viktig for aktørene at koblingen ikke førte til utgifter og merarbeid for dem. Det var avgjørende for dem at PAHN stod for midlene til gjennomføringen.
- Vi har fått bekreftet at Scenewebdatabasen teknisk sett er fleksible nok til å kunne ta imot data fra flere systemer.
- Det tekniske arbeidet med å koble databasene sammen er en engangsjobb, men koblingen av at importene krever manuell kvalitetssjekk for å unngå feil og duplikater. Dette arbeidet vil kreve fortløpende ressurser (arbeidskraft).
- Vi ser at tross effektiviseringen denne metodikken medfører så kreves også, som nevnt, en del manuelt arbeid for oss i Sceneweb. Både i dialogen før og under inngåelse av kontrakt, dialog med IT-utviklerne underveis i prosessen, testing av løsningene, og manuell godkjenning av importene i etterkant.

- Det er ingen metadata på filene, så disse krever ekstra arbeid for å matche vår metadatastandard. Her vil vi i videre arbeid (neste steg av prosjektet) se på mulighetene for å generere noe metadata automatisk. Vi vil videre vurdere mulighetene for å generelt produsere mer metadata automatisk enn det som kommer inn i dag, slik at vi minsker det manuelle etterarbeidet så mye som mulig.

Vi ser stort potensiale i metodikken vi har utviklet i denne piloten. Metadataene produseres kun en gang, og i tillegg til metadataene vi vanligvis registrerer på en produksjon får vi inn fullstendige spille-/turnélistene, og i to av databasene får vi også inn bilder, plakater og/eller video. Tross det manuelle arbeidet som kreves for å kvalitetssjekke importene så er dette en tids- og kostnadseffektiv metode fremfor om vi skulle gjort alt arbeidet manuelt. Per dags dato mener vi det er hensiktsmessig å videreføre denne metodikken. Når vi har mer erfaring med dette vil vi ta en ny evaluering og gjøre en ny vurdering om hvorvidt den manuelle kvalitetssikringen er nødvendig. De gode resultatene fra piloten førte til at vi i 2025 har startet en videreutvikling av arbeidet til et 3-årig prosjekt.

Forklaring på endringer i tidsplanen

Prosjektets sluttdato ble utsatt fra 15.09.2024 til 31.12.2024. Det var flere årsaker til dette: Det tok lenger tid enn forventet med programmering av de to første koblingene; mellom Scenewebdatabasen og kalenderdatabasen til Scenekunst.no og repertoardatabasen til Riksteatret. Dialogen med noen av teatrene tok også lenger tid enn forventet. Dette hadde flere årsaker; 1) interne ressurser i PAHN - arbeidet med piloten kom på toppen av en allerede stor arbeidsmengde, 2) flere teatre trengte god tid til å forstå hva deling av data innebærer og for å tenke seg om og komme frem til hvilke data de ønsket å dele med oss.

Forklaring på endringer i antall databaser vi koblet opp

I utgangspunktet var planen at vi skulle koble 7 databaser til Sceneweb. Det er to årsaker til at vi endte med å koble opp 6, altså 1 mindre enn planlagt:

- 1) Kostnadene knyttet til importen av kalenderdatabasen til scenekunst.no ble NOK 7500,- høyere enn antatt.
- 2) Tildelingen fra Arkivverket var NOK 26 344,- lavere enn søknadssum.

Vi kuttet ett av teatrenes database. Ett av de 6 teatrene vi var i kontakt med er i en veldig tidlig fase i bruken av sin database. De var i utgangspunktet positive til deling av data, men ønsket selv å avvente en kobling til Sceneweb til de selv har mer erfaring med planverktøyet. Dette var en god løsning for oss også, da det var viktig å gjennomføre koblingen til scenekunst.no i og med at denne er helt forskjellig fra teatrenes databaser. Vi ønsker å arbeide videre med oppkobling til flere forskjellige databaser i fremtiden og da var det viktige erfaringer for oss å teste metoden på forskjellige databasesystemer.