Structured Data for Performing Arts History: An Introduction to a Special Issue of Data Papers

Arts and Media

Thunnis van Oort
University of Amsterdam, Amsterdam, Netherlands
t.vanoort@uva.nl

Julia Noordegraaf
University of Amsterdam, Amsterdam, Netherlands
j.j.noordegraaf@uva.nl

Abstract

This introduction to the special issue of performing arts data papers starts off by dwelling on the purpose of the (re-)use of structured data for performing arts history, and on the benefits of publishing data papers about these data sets. The authors present several observations on the general characteristics of data models in this scholarly domain and conclude by discussing each of the contributions to this special issue of the Research Data Journal for the Humanities and Social Sciences.

Keywords

data papers – data models – performing arts – history

Cultural performances in theatre, music and film have contributed vividly to the formation of individual and social identities in Europe. The physical space of the auditorium is charged with social and cultural meanings, bringing together individuals from all walks of life and at the same time demarcating their social differences (Balme, 2014; Furnée, 2012). To grasp these processes of identity formation we can study the moment of cultural exchange, as it takes place on the stage or screen as well as in the minds and bodies of the spectators and...

Traditionally, scholarship in music, theatre and film history has prioritised the study of the production of the artwork over its consumption. But if we want to evaluate and understand the social and political significance of the performing arts, an analysis should not be limited to the artwork itself, but needs to include the audience as part of the wider framework that situates the performing arts in society. Already from the 1980s, scholars have complemented the prevalent research on the form, style and performance of theatrical, musical and cinematic works of art with research that investigates the distribution and reception of performing arts (Allen & Gomery, 1985; Booth, 1991; Fischer-Lichte, 1997; Garncarz, 2015; Gerhard, 1992; Gurr, 2004; Johnson, 1995; Biltereyst, Maltby, & Meers, 2011; Sedgwick, 2000; Staiger, 1992; Weber, 1975; Wollenberg & McVeigh 2004).

The sheer volume of data in the digital age and the potential of its connectedness to other (types of) data has created new opportunities for historical research and has given a boost to distribution, reception and consumption studies (Acland & Hoyt, 2016; Leonhardt, in press). In particular, such datasets provide the empirical basis for research into the distribution and consumption of performing arts, by listing repertoires and programming as well as the venues, persons and organisations involved in performing arts events, revealing networks and patterns hidden before from the historian’s view (Nijboer & Rasterhoff, 2018).

This themed issue bundles nine data papers that engage in the writing of performing arts history. We hope, and expect, that this is just a starting signal for more to come.

One important reason to publish data papers is that they provide more transparency in scholarship and can, thus, help to improve the accountability of research on the history of the performing arts. Limitations of space in traditional academic formats such as the journal article do not always allow for a full account of the datasets used in the research. Data papers can provide detailed documentation and acknowledgements on the formation of the data set without interfering with the argumentation structure commanded by a regular research article. Data papers also make it easier to give citable credit due to those collaborators who helped create the data set, but who did not necessarily contribute to the writing of the research output. The publication of datasets and accompanying data papers such as the ones included in this issue allow other scholars to test the findings and, perhaps, to replicate the study. Also, the publication of data sets with accompanying data papers can be a remedy for
the problem of “orphaned” research databases, abandoned after project finances ran out.

One specific form of data re-use would attempt to cut across traditional disciplinary borders and aim for a more holistic view of the consumption of music, theatre and film, because for audiences, those boundaries, of course, did not follow academic institutional partitions. Some of the papers already hint towards such a transdisciplinary approach, such as the Antwerp Zoo collection that encompasses both music and theatre (Engelen, Crombez & Vande Winkel, 2020), or the Opera database that targets an inherently cross-medial phenomenon (van Nieuwkerk, Salters, Helmers, & Kisjes, 2020).

The combination of existing and new datasets also enhances the opportunities for comparisons between cities or countries. The performing arts have a fundamentally international character, travelling across national boundaries and language divides. To properly study them means looking beyond local and national borders (Biltereyst, van Oort, & Meers, 2019). With a growing set of (well-described) datasets available, it will become increasingly feasible to do just that (van Oort & Noordegraaf, in press). And the promise of Linked Open Data as a format optimised for connecting data sets will only increase the potential of doing performing arts history with the help of structured data.

Presently, in Europe alone, dozens of databases containing historical information on performing arts are available online. In 2017, an inventory of European databases in the domains of theatre, cinema and music history was compiled, and their underlying ontologies were analysed (Baptist, 2017; Baptist, van Oort & Noordegraaf, in press). A shared characteristic of these databases is that they are governed by the basic ontological hierarchy of the “work” (“concerns the elements that never change about a work”), “production” (“concerns the team that put the work on stage”) and “performance” (“concerns the people on stage or in the orchestra pit who perform a certain production of the work on a particular day”) (Royal Opera House, n.d.).

The data sets in this inventory are based on one of two main types of data models: item-oriented or event-oriented. Item-oriented data models are structured around specific items such as venues (buildings, locations), works (plays, librettos) or persons (playwrights, musicians). Venue-oriented data models may take one or multiple venues as a starting point. An example is the Royal Opera House Performance Database, which centres around one particular institution as an entry point to its performances. The Theatrescapes Research Tool, by contrast, targets a multitude of venues as places of performance (Leonhardt, n.d.). Other item-oriented databases take as a starting point the
“work” written or produced by a certain playwright or composer. For example, the Out of the Wings or the PLAYOGRAPHYIreland projects foreground the contextualisation of adapted or translated stage plays, to which data on the production and performance history are then linked (Irish Theatre Institute, n.d.; Out of the Wings, n.d.). Another example of a “work”-oriented database is the Staging Beckett Project, which starts from the works of the Irish playwright as an entry point for exploring the professional productions of his plays throughout the United Kingdom and Ireland. A final group of databases that use an item-based data model are thematic ones, such as the National Theatre’s Black Plays Archive or the (now-defunct) Exile Remains Database on Spanish Republican exile theatre – the fact that this resource can no longer be found online only emphasises the need for data archiving that is addressed in this special issue and in this journal more generally (Exile Remains Database, n.d.; National Theatre, n.d.).

In contrast to item-oriented data models, the event-oriented data model takes a combination of items as its basic structuring principle. The creators of the Australian theatre database AusStage define an event as “a distinct happening defined by title, date and venue,” indicating how an event is usually at least made up of a date, a place and a work (Ausstage, n.d.). In this sense, event-oriented data models combine various elements of item-oriented data models, linking basic entities like “work,” “person,” “venue,” and “company”, which together constitute the performance as an event. The datasets presented in this special issue are all event-oriented, except for the Dutch Theatre Production database, where productions rather than performance events are central.

Event-driven databases lend themselves especially well to comparative research, across local and national borders as well as across different performing arts disciplines (Biltereyst & Meers, 2016; Nijboer & Rasterhoff, 2018; Noordegraaf, Lotze & Boter, 2018). Linking and comparing different historical data sets does require a reflection and mutual agreement on the definitions of the objects of study. Taking the performing arts event as focal point for investigating historical audiences, how should its basic characteristics be defined? What are significant similarities and differences between theatrical, musical and cinema performances and to what degree can we use shared ontologies? And which type and structuring of information would need to be included in event-based ontologies and taxonomies for the performing arts? (van Asseldonk, van Mensch & van Vliet, 2009).

This special issue can help address such questions by bringing together both recent contributions to a “databased” approach to performing arts
history with the work of trailblazers who started to collect structured data in the 1970s and 1980s when computer facilities began to become accessible to historians. The selection of nine data papers covers various performing arts disciplines, time periods and localities and describes data sets that range widely in scale and complexity. All described data sets are deposited in trusted repositories and are publicly available via the links and identifiers provided.

As early digital humanists, Simon McVeigh, Henk Gras and Karel Dibbets demonstrated the potential of structured data and computational techniques to produce new insights about the audiences and economics of the performing arts. A pioneer of the social history of the theatre audience, the Dutch historian Henk Gras has adopted statistical methods to analyse patterns in the social composition of theatre audiences. In the early 1990s, when relational database software such as Microsoft Access became available to users without specialised knowledge in information technology, he started compiling a massive relational database connecting theatre-goers (mainly subscribers) to their demographic and social status in Rotterdam during the long nineteenth century. The theatre historiographical “master narrative” has long claimed that the emergence of melodrama in the mid-19th century was caused by an influx of lower-middle-class and even working-class audiences that drove the upper strata of the social spectrum away from drama to the opera. By tracking down the identity of 19th-century theatre subscribers in Rotterdam, then matching these names with class markers such as occupation (from address books) and income (from tax records), Gras et al. (2011) could establish that the melodrama audiences in Rotterdam were of solidly middle- and upper-class status. In his contribution to this theme issue, Van Vliet describes the genesis of Gras’ dataset, inviting theatre and social historians to explore its rich and as yet unexplored opportunities (van Vliet, 2020).

Another “early adopter” that recognised the potential of tabular data for researching cultural history is the British musicologist Simon McVeigh. In the 1970s, he started building a concert database that allowed him to trace the rise of London’s public music culture from its dawn in the 1670s until its fruition around 1800 as one of the world’s most prominent musical ecosystems (McVeigh, 2012). In the data paper, McVeigh (2020) describes in full detail which choices were made in its design and how the data set was compiled. He argues why its highly systematic and comprehensive nature still provides significant advantages over the current possibilities for searching London newspapers digitised and OCR’ed long after the dataset was compiled. Furthermore, in comparing his dataset to those documenting theatre or cinema performances, McVeigh points out that
the variable structure of concert programmes sets them apart from theatrical performances: some concerts may have 20 items, others just one; a single item may have seven instruments and performers listed, or none; the same symphony or aria may be known under a variety of names in different languages, and so on.

McVeigh, 2020, p. 59

A third groundbreaker, who applied early digital methods to investigate the economic underpinnings of the Dutch cinema culture, is the Dutch economic historian and film scholar Karel Dibbets (1947–2017). During the research for his MA thesis on chain formation in the Dutch post-WWII cinema business in the 1970s, he had collected a database with information on cinema exhibitors and distributors in the Netherlands. In collaboration with a programmer at the faculty of Social and Behavioural Sciences and using the University of Amsterdam’s then sole mainframe computer, Dibbets used network analysis to make sense of the data. The analyses demonstrated how ownership structures in the Dutch cinema economy had gradually become concentrated during the post-war period (Dibbets, 1980). This research in the late 1970s was a starting point for a much more ambitious project called Cinema Context, an online encyclopaedia of Dutch cinema culture that to this day is being managed and developed at the University of Amsterdam. Currently, a project is running that will make the Cinema Context data available in RDF format (van Oort, n.d.-a). In their data paper, Thunnis van Oort and Julia Noordegraaf recount the creation history of Cinema Context and describe the data model in more detail, providing a critical guide for users of the data set (van Oort & Noordegraaf, 2020).

The Cinema Context data model has been used as an inspiration for similar projects abroad, such as the Flemish Cinecos project (FWO, 2018–2021) and the comparative project European Cinema Audiences (AHRC, 2018–2021). Agreeing on basic principles of how to organise this film-related data opens up prospects for transnational comparative research. The data paper “Mapping film programming across post-war Europe (1952)” presents an example of such a transnational comparative initiative in practice (van Oort et al., 2020). This research collaboration that involves a dozen scholars brings together film programming data from several European cities. As comparisons of movie-going patterns between European countries are still rare, this paper offers a model for constructing a data set which can be replicated, scaled up, and used to compare, contextualise, and eventually construct theories about practices of cinema-going across countries at a global level. This case study highlights some of the challenges that arise in transnational, collaborative data collection, which
occur on several levels, such as the selection of the case studies and the definition of and criteria for basic units of comparison.

A fourth veteran data set described in this data paper collection even dates back to the 1950s. In her contribution to this issue, Floortje Bakkeren provides an in-depth account of how the Dutch Theatre Production Database has developed from its inception in 1952 until the present (Bakkeren, 2020). Standing out from the other data collections discussed here, the Dutch Theatre Production Database did not originate as a research database initiated for scholarly purposes. Instead, it initially functioned as the central information system of a theatre heritage institution, first known as Toneelmuseum (“Stage Museum”), later Theater Instituut Nederland (“Dutch Theatre Institute”), and now part of the Special Collections of Amsterdam University Library. By retracing the history of the database and how the protocols for data collection and entry changed, the paper offers users much more precise insights into the range and representativeness of the data presented, increasing its usability for research (see for a recent example of the application of the data: van Oort & Kisjes, 2019).

A database that partly overlaps with and also significantly adds to the Dutch Theatre Production Database focuses on operatic productions in the Netherlands. Mascha van Nieuwkerk and her colleagues in their contribution on this dataset recount the process of digitising a printed collection titled *Annalen van de Nederlandse Opera-gezelschappen* (“Annals of the Opera Companies in the Netherlands,” 1996), converting it to a relational database (van Nieuwkerk et al., 2020). The original database in book form lacked acknowledgement of the criteria and sources used to systematically collect information on over a century of operatic productions in the Netherlands. A comparison with the Dutch Theatre Production Database in the paper helps to shed more light on the coverage of the opera database. Van Nieuwkerk et al. demonstrates how the tricky question of defining opera complicates ambitions of completeness. Interestingly, the data model developed for this project has triggered the interest of the Dutch National Opera, that is currently restructuring its internal information system (van Oort, n.d.-b).

Where the previously described datasets have a national or urban scope, the remaining databases described in this collection focus on the performances at a single institution. Leen Engelen, Thomas Crombez and Roel Vande Winkel present a data set of film screenings and concerts that have taken place in the Antwerp Zoo, which boasted a theatre that could seat over 2,000 patrons. The Antwerp City Archive holds a remarkable collection of printed programmes, that has formed the foundation of the database. Scans of all the programmes are also available in conjunction with the database on the website.
van Oort and Noordegraaf

In their paper, Engelen, Crombez, & Vande Winkel (2020) describe how they transformed this collection into a structured database, which additional sources they used, and which decisions they had to take when designing the data model, concluding with an evaluation of the scope of the data set.

Similarly, an archived collection of programmes prompted the construction of a database of nineteenth-century concert programming of the prestigious Felix Meritis concert hall in Amsterdam – the predecessor of the Royal Concertgebouw before the latter opened in 1888. The felix database underlines McVeigh’s point that the variable structure of music programming creates specific challenges for the data model. The Felix Meritis programming data demonstrates the complexity of nineteenth-century musical programming, which often consisted of a wide range of short pieces that are difficult to identify or classify. In this data paper, Mascha van Nieuwkerk, Harm Nijboer and Ivan Kisjes (2020) explain their decision to keep the original, ad verbatim descriptions of the concert items. The way the individual items were clustered into full concert programmes and the period-specific genre designations both offer rich material for historical analysis.

Finally, the onstage database contains information on programming and box office figures for Amsterdam’s prominent municipal theatre ‘De Schouwburg’, dating back to its foundation in 1638. This dataset, described in detail in the paper by Blom, Nijboer and van der Zalm (2020), is an apt illustration of how structured data on theatrical performances can lead to surprising revisions of established canons. For generations, Dutch school children have been taught that Joost van den Vondel was the most prominent Dutch playwright of the 17th century. A recent analysis of theatre programming in Amsterdam during the Golden Age, however, demonstrated that Vondel was surpassed in popularity by translated Spanish playwrights such as Lope de Vega, which have now vanished from the Dutch collective memory (Blom & van Marion, 2017). Studying Dutch theatre history from the consumption perspective allows us to understand that Dutch audiences were much more internationally oriented than subsequent nationalist historiography had made us believe.

We like to thank all the authors, reviewers and the journal editors for their contributions to this collection.

References


Exile Remains Database (n.d.). No longer retrievable via http://www.hispanicexile.bham.ac.uk/.


Nijboer, H., & Rasterhoff, C. (2018). Linked cultural events: Digitizing past events and implications for analyzing the ‘Creative City’. In S. Münster, K. Friedrichs,


