

A Graph Database for Performing Arts in Switzerland

Birk Weiberg,
Project Manager, Foundation SAPA,
Swiss Archive of the Performing Arts

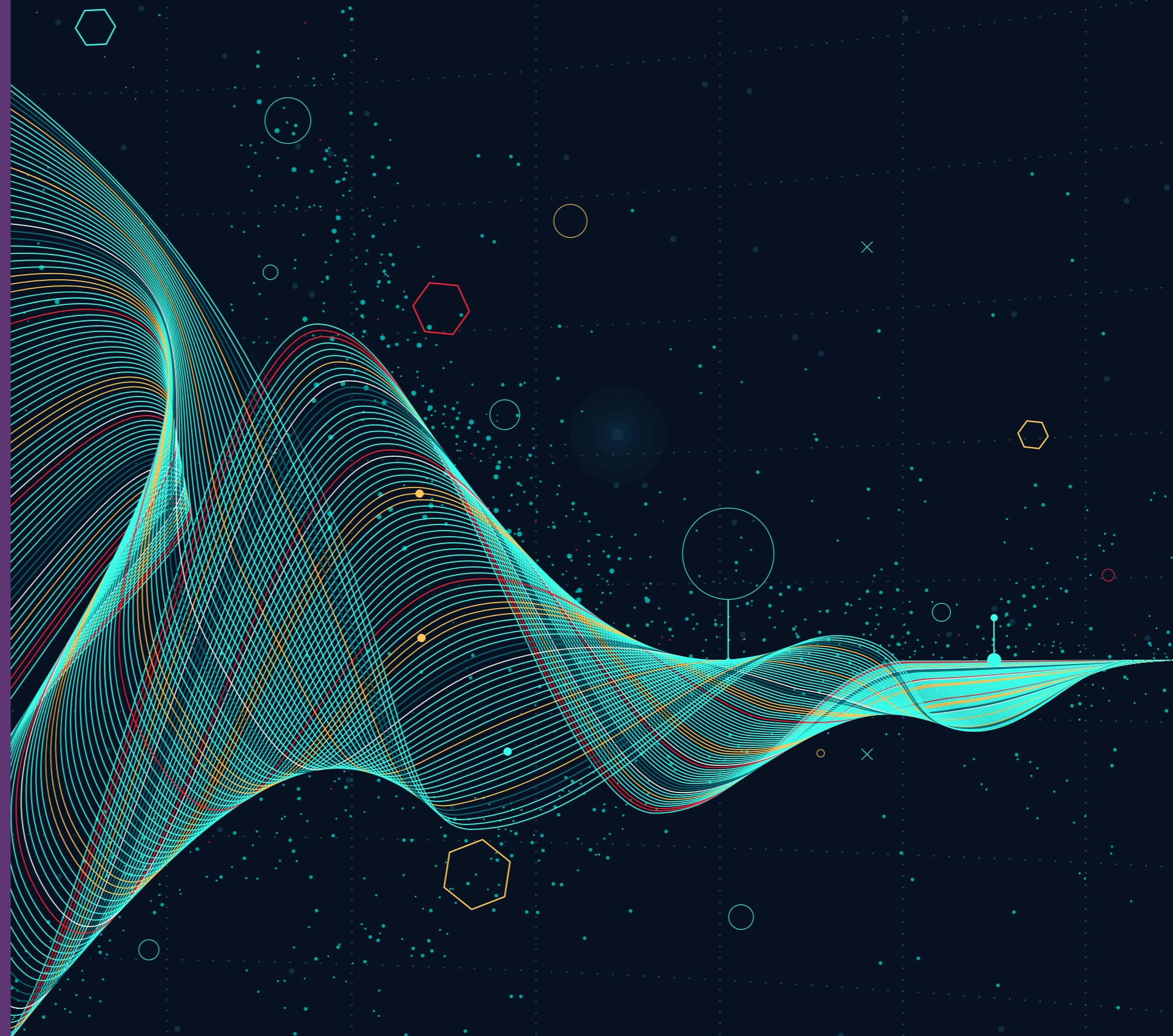
The Swiss Archive of the Performing Arts (SAPA) emerged in 2017 from the joining of three independent archives in Bern, Lausanne, and Zurich. The histories of SAPA's predecessors demonstrate the ever-developing approaches in documenting performing arts—a field that, like ICH in general, cannot be preserved in its original form. In our case, the documentary efforts started nearly 100 years ago with an initiative to collect books and other documents on theater culture in Switzerland. Later, objects such as stage design drawings and models followed and found their way into a small museum. The increasing usage of video in contemporary dance in the 1990s resulted in the establishment of the former Dance Media Library in Zurich in 2005, the youngest of the three SAPA predecessors.

In this short chronology of evolving documentation practices, the establishment of SAPA as a joint cultural heritage institution for theater and dance in Switzerland falls together with an increased interest in collecting and combining different kinds of data as a more comprehensive mode of documentation. A central project of the persisting three sites' ongoing integration is the merger of various legacy databases into one shared graph database: performing-arts.ch.

SAPA's predecessor institutions followed distinct approaches in their usage of databases. In one case, a traditional archival database represented the holdings hierarchically according to their provenance. In another case, there were various specialized databases with custom fields to store domain-specific information. The information within these databases can be divided in two categories: some describe documents or artifacts in our holdings that are related to works or protagonists of performing arts while others describe works and protagonists of the field even if we do not have any archival material. Among the latter type of information are production details and credits of more than 60,000 productions shown in Switzerland since the late nineteenth century. This documentary part of the database also serves as an authority file that provides stable identifiers and basic facts about more than 30,000 persons, groups, and venues.

Our efforts to connect these different kinds of information coincided with a broader movement within the archival community to question the prevalence of the principle of provenance that preserves information regarding the origins of a collection and makes it the primary way of accessing it. The growing skepticism with this perpetuation of established narratives and also power structures utilizing archival organization is expressed best in the ongoing development of the new descriptive standard Records-in-Context (RiC) by the International Council on Archives (ICA) as a replacement for its current standard ISAD(G). As the name suggests, RiC will allow the identification of not only records and their creators but also subjects and other relevant properties by using unique identifiers.¹

While this marks a significant step, RiC remains an archival standard that does not provide the means to represent domain-specific information. Thus, we combined it with the CIDOC CRM standard developed for museums, and its extension FRBRoo for bibliographic information. The FRBRoo ontology in particular has proven to be helpful as it provides classes for performing arts



Big data visualization digital © max776, kr.123rf.com

that distinguish between concepts, productions, and actual performances.² What makes both CIDOC CRM and FRBRoo particularly suitable for ICH data is that they are event-centric—i.e., they conceive objects through the activities that have produced and changed them, whereas older data models have been restrained to the description of the physical objects in the custody of museums or archives.

The development toward data models with higher complexity and the expectation of a more detailed and thus realistic rendering of performing arts was also supported by the development of graph databases. Where the still-predominant relational databases are structured as a set of tables with highly uniform entries, graph databases contain data as virtually unlimited networks of nodes and edges. They are only delimited semantically by an ontology. As intended by the creators of RiC, this rhizomatic structure provides new perspectives on the data that overcome the one of provenance. The inherent multidimensionality of knowledge graphs can also help to transcend anthropocentrism—in a database where virtually everything can be identified and thus linked to its other occurrences, anything (e.g., a venue) can become the center of the network.

The migration of the legacy databases has been a slow and diligent process. This started with developing a data model that can describe the different artifacts and also contain documentary data regarding casting and production structures.³

Information that so far was stored in free text fields needed to be cleaned and formalized to be able to reconcile the denoted entities. Finally, an infrastructure was needed to maintain and edit the data. While the first step simply asks for time and perseverance, the second requires a software that ideally stores Resource Description Framework (RDF) data natively and allows custom interfaces to be built according to the data model. Our database is built upon the *metaphactory* knowledge graph platform, a software that has become increasingly popular mainly for industry projects that work with complex data stored in triplestore databases but can also easily be used for cultural data.

The choice of a graph database with persistent identifiers for all relevant entities and a platform-independent ontology is relevant for a sustainable data management that follows the FAIR data principles. Entities such as persons, groups, or places, which are also described by other authority files, are widely connected with these, and the same applies to controlled vocabularies. To improve the findability, we have registered many entities with Wikidata, the structured data platform by Wikimedia, which is becoming increasingly relevant as a hub for connecting cultural heritage institutions. Missing entries are added to Wikidata and then can be enriched by other users or bots. Our users can retrieve the data they need from our database through a SPARQL-interface and use it under a CC BY-SA license. Interoperability of the data at the moment primarily means for us that we can provide high-quality data to aggregators like Memoriav, the Swiss network for the preservation of national audiovisual cultural heritage and its database memobase.ch, and the database performing-arts.eu, run by the Specialised Information Service (SIS) Performing Arts for Theatre and Dance Studies, a key information source for the German-speaking research community.

performing-arts.ch is an ongoing project, and we will continue the increase of the quality of the data by identifying further protagonists, ingesting further legacy databases, and thinking about alternative interfaces that the knowledge graph approach can facilitate. ■

Title	Premiere	Producer	Venue
D.S.C.H	1994-10-08	Compagnie Linga	L'Octogone Théâtre de Pully
Petites approches sentimentales	1997	Drift Compagnie	
Zugollos, Forum 1	1996-07-04		
Couples	1988-11-27	Compagnie 100% Acrylique	La Traverse
Living Games	1987-06-12	Ballet Junior	
two, too	2001-09-04	anna+huber+compagnie	Proberaum Luzern
Deus ex machina	1997-11-21	Ventura Dance Company	Schauspiel-Akademie Zürich
Les buveurs de brouillard	1999-12-31	Die Buissonnière	
Tránsito flamenco	2002-11-30	Flamencos en route	Kurtheater Baden
SPP ET ODC (Suite pour percussion et quatuor de danse contemporaine)	1997-11-12	Compagnie Linga	Salle Métropole

Screen capture.
© performing-arts.ch.

NOTES

1. Bogdan-Florin Popovici, “A Broader Perspective on Records as Seen by Records-in-Contexts,” *Comma* 2016:1–2 (January 2018): 189–98, doi:10.3828/comma.2016.19.
2. Martin Doerr, Patrick Le Boeuf, and Chryssoula Bekiari, “FRBRoo, a Conceptual Model for Performing Arts,” in *2008 Annual Conference of CIDOC, Athens, 2008*, 15–18.
3. For the original data model, see Beat Estermann and Christian Schneeberger, “Data Model for the Swiss Performing Arts Platform” (Draft Version 0.51, 2017), available at <https://datahub.io/dataset/spa-data>. The current implementation is documented at <https://sapa.github.io/spa-specifications/>.