

musicSpace MS workshop paper

ABSTRACT

musicSpace: Improving Access to Musicological Data

mc schraefel, David Bretherton, Daniel Alexander Smith, and Joe Lambert

Efforts over the past decade to digitize scholarly musicological materials has revolutionized the research process, however online research in musicology is now held back by the segregation of data into a plethora of discrete and disparate databases, and the use of legacy or ad hoc metadata specifications that are unsuited to modern demands. Many real-world musicological research questions are rendered effectively intractable because there is insufficient metadata or metadata granularity, and a lack of data source integration. The "musicSpace" project has taken a dual approach to solving this problem: designing back-end services to integrate (and where necessary surface) available (meta)data for exploratory search from musicology's key online data providers; and providing a front-end interface, based on the "mSpace" faceted browser, to support rich exploratory search interaction.

We unify our partners' data using a multi-level metadata hierarchy and a common ontology. By using RDF for this, we make use of the many benefits of Semantic Web technologies, such as the facility to create multiple files of RDF at different times and using different tools, assert them into a single graph of a knowledge base, and query all of the asserted files as a whole. In many cases we were able to directly map a record field from a partner's dataset to our combined type hierarchy, but in other cases some light syntactic and/or semantic analysis needed to be performed. This small amount of work in the pre-processing stage adds granularity that significantly enriches the data, allowing for more refined filtering and browsing of records via the search UI. Significantly, although all the data we extract is present in the original records, much of it is neither exposed to nor exploitable by the end-user via our data providers' existing UIs. In musicSpace, however, all data surfaced can be used by the musicologist for the purposes of querying the dataset, and can thus aid the process of knowledge discovery and creation.

Our work offers an effective generalizable framework for data integration and exploration that is well suited for Arts and Humanities data. Our benchmarks have been (1) to make tractable previously intractable queries, and thereby (2) to accelerate knowledge discovery.