# UNIFIED ACCESS TO MEDIA INDUSTRY AND ACADEMIC DATASETS: A CASE STUDY IN EARLY MUSIC

**David M. Weigl, Kevin R. Page**
Oxford e-Research Centre
Univeristy of Oxford
{first.last}@oerc.ox.ac.uk

**David Lewis, Tim Crawford**
Department of Computing
Goldsmiths, University of London
{f.last}@gold.ac.uk

**Ian Knopke**
British Broadcasting
Corporation
ian.knopke@bbc.co.uk

## ABSTRACT

Within the Fusing Audio & Semantic Technologies (FAST) project, we are linking and providing access to datasets from industry and academia, here with a focus on early European music. We have developed SALT, a tool for the semantic alignment and linking of datasets describing shared instances (e.g. composers, performers, songs) but lacking common identifiers; and an augmented web resource for the BBC Radio 3 Early Music Show, enabling views of the underlying unified data. The constituent datasets, the linking structure created by an expert musicologist using SALT, and provenance regarding the musicologist's matching decisions are all expressed as RDF triples, accessible through a public SPARQL endpoint exposing the semantic structure that also drives the augmented web resource and the alignment tool. By publishing this shared structure as Linked Data, we expect further value to accrue as reuse of the data in other contexts is facilitated.

## 1. INTRODUCTION

Manual alignment of humanities datasets is time consuming and prone to error; in lieu of common identifiers, consideration of non-identifying surface labels and metadata results in a combinatorial explosion where all potential entity pairings represent match candidates to be considered. Fully automating the process is difficult due to ambiguities inherent in the data, e.g. where shared textual labels refer to different entities, or where different sources present divergent information relating to the same entity. In previous work [2] we have presented our Semantic Alignment and Linking Tool (SALT) that addresses these problems by putting a human expert in the driver's seat while minimising the "drudge work" of the alignment process. Here, we present an augmented version of the Early Music Show website, building on an initial application of SALT by a musicologist with domain expertise in early music to combine music library resources and BBC programme data.

## 2. AUGMENTED EARLY MUSIC SHOW WEBSITE

The Semantic Linking of BBC Radio (SLoBR) tools combine a number of constituent datasets including Early Music Show broadcast data published by the BBC; the SLICK-MEM dataset [1] comprising digitised music library resources (high-resolution musical score images and associated catalogue metadata) from the British Library and other sources; and further external sources of biographic and bibliographic metadata, including DBPedia, a publisher of structured information extracted from Wikipedia; Linked-Brainz, a service providing Linked Data access to MusicBrainz; and VIAF, the name authority service. The graph structure underlying this data is expressed as RDF triples, allowing the constituent datasets to each employ different data schemas without hindering interoperability.

The Early Music Show programme website currently supports navigation based on episode broadcast date. By virtue of the additional data and linking structure made available through our project, we have augmented this resource to enable further navigation vectors focusing on composers, performers, and works. This enables novel traversals of the programme data based on particular agents or works of interest, and on associated contextual information. We can thus link, e.g., from a page describing a particular musical work with associated digitised score, to a listing of radio episodes featuring compositions by contemporaries of the work's composer. This serves as an example of how Linked Data techniques and publicly available academic datasets can be used to improve data resources in a media industry setting, providing new experiences to the consumer while facilitating scholarly research.

## 3. REFERENCES

[1] Tim Crawford, Ben Fields, David Lewis, and Kevin Page. Explorations in Linked Data practice for early music corpora. In *JCDL 2014*, pages 309–312. IEEE, 2014.

[2] D. M. Weigl, D. Lewis, T. Crawford, and K. R. Page. Expert-guided semantic linking of music-library metadata for study and reuse. In *DLfM 2015*, pages 9–16. ACM, 2015.