

A Shared Open Vocabulary for Audio Research and Retrieval

György Fazekas Alo Allik Simon Dixon Mark Sandler
Centre for Digital Music, Queen Mary University of London

The Shared Open Vocabulary for Audio Research and Retrieval project investigates how audio research communities would benefit from using shared open vocabularies with a focus on the Music Information Retrieval (MIR) community. Researchers in speech, music, bioacoustics or environmental audio signal processing and retrieval increasingly use common sets of features to characterise audio material, while large data sets are released for public and scientific use. The development of data sets and research tools however are not governed by shared open vocabularies and common data structures. While there has been tremendous work to create easy to use feature extraction tools, it remains difficult to assess for instance, whether audio features computed by different tools are mutually compatible or interchangeable. Similar problems arise with the release of data sets like the Million Song Dataset [1] or the SALAMI dataset [3], in a variety of different formats, as well as in the use of various Web APIs.

The problem affects several communities including audio signal processing and MIR researchers, as well as audio archives, libraries, broadcasters and creative industries that may utilise the outcome of the above research activities. In this project we investigate (1) if and how audio research communities would benefit from using interoperable file formats, data structures, vocabularies or ontologies, (2) what are the primary needs of MIR researchers, and (3) what are the main barriers to the uptake of shared vocabularies.

The existing audio features ontology was created within the framework the Music Ontology, which was developed during the Online Music Recognition and Searching 2 (OMRAS2) [2] project to solve a more general problem regarding interoperability between music related data sources. While the Music Ontology has been widely adopted by researchers and user communities, Semantic Web programmers, as well as the industry, the adoption of the audio features ontology remains relatively limited, primarily due to limited exposure to research communities, fuzzy domain boundaries and incomplete vocabulary with regards to user needs.

In this talk we discuss the benefits of shared vocabularies and common data formats and the ways in which Semantic Media technologies can facilitate interoperability and sustainability of research workflows with the emphasis on audio research. We reflect on what the project has achieved in its beginning phase, which includes cataloguing audio features and reviewing existing research tools and vocabularies, as well as engaging research communities in the decision making process. The SOVARR project is focused on community involvement and we aim to directly engage experts in the audio research and linked data communities in the process of developing shared vocabularies and research tools.

Further we examine the main obstacles to the creation of a single shared vocabulary and consider more feasible alternatives such as a framework of several shared modules that are more domain or task specific. A more fine tuned approach that takes into consideration the characteristics of a specific tool or a part of the community with specific research practices could significantly increase the adaptability of shared vocabularies. These modules could potentially be linked on a higher level or be rooted in a dynamic, community authored vocabulary. Semantic Web ontologies have properties that are ideal to support our requirements. It is possible to create flexible and modular systems that still support the unique identification of terms, the possibility of establishing hierarchical or equivalence relationships between them, and describing the meaning of data at different levels of detail.

References

- [1] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 591–596, Miami, FL, USA, 2011.
- [2] György Fazekas, Yves Raimond, Kurt Jakobson, and Mark Sandler. An overview of semantic web activities in the OMRAS2 project. *Journal of New Music Research (JNMR)*, 39(4), 2010.
- [3] Jordan Bennett Louis Smith, John Ashley Burgoyne, Ichiro Fujinaga, David De Roure, and J. Stephen Downie. Design and creation of a large-scale database of structural annotations. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 555–560, Miami, FL, USA, 2011.