SHARED OPEN VOCABULARIES AND SEMANTIC MEDIA

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

ABSTRACT

This paper presents two ongoing projects at the Centre for Digital Music, Queen Mary University of London. Both projects are investigating the benefits of common data representations when dealing with large collections of media. The Semantic Media project aims at establishing an open interdisciplinary research network with the goal of creating highly innovative media navigation tools, while the Shared Open Vocabulary for Audio Research and Retrieval (SOVARR) project builds on community involvement to improve existing tools and ontologies for MIR research. Common goals include bringing together experts with various research backgrounds and establishing open vocabularies in combination with semantic media technologies as viable tools for sustainable and interoperable workflows. In this paper, we summarise our projects as well as the results of the Shared Open Vocabularies session that took place at ISMIR 2012.

1. THE SEMANTIC MEDIA PROJECT

The profusion of digital content now available to an average consumer is overwhelming, potentially forcing consumers into increasingly narrow bands of media experience as they retreat to limiting their choices as a coping strategy. Professional recommenders, such as newspaper film and TV reviewers are similarly overwhelmed, and paradoxically, more relied upon by consumers whilst considered less relevant, as automated and semi-automated recommendation systems emerge for movies (e.g. The Netflix competitions), music (e.g. last.fm), or books (e.g. Amazon).

The Semantic Media Network project addresses the challenge of time-based navigation in large collections of media documents. The project focuses on investigating new ways to empower users to find relevant content, and exploring how industry and universities can work together in this field. In particular, one of the project's central ideas is that media annotation should occur within the production process, so that not only consumers benefit from the introduction of new search, browsing, and recommendation technologies, but also content producers like composers, musicians, script-writers, directors, or actors. Furthermore, annotating content as early as possible allows for integrating knowledge of the production workflow, which leads to simplified and hence more robust automatic procedures, as well as more detailed metadata and richer user interfaces. Additionally, managing and exposing this metadata using modern semantic web and linked data technology allows for uniting various sources of information, which enables users to more effectively identify relevant content and thus helps to widen the consumer's increasingly narrow bands of media experience. The project's scope is the whole life-cycle of content with the goal of empowering human producers and consumers to effectively reuse, re-purpose, and personalise material, whether for entertainment, news, documentaries, education, interviews, health-care, science or security.

In this context, the projects aims at bringing together a critical mass of skilled people from across the ICT sector: academic researchers and commercial partners (semantic web, knowledge engineering, databases, machine learning, text mining; human interactivity and social computing; signal processing, image, audio and video analysis) as well as media specialists and practitioners from the creative industries. Effective communication and collaboration across many different fields require a common language. Therefore, defining shared vocabularies and ontologies capturing knowledge about media data is not only essential for the development of new methods and tools as stated above, but also enables researchers to truly collaborate as it becomes straightforward to integrate and unify results across all fields. For this reason, the SOVARR project is of central importance for the Semantic Media project.

2. THE SOVARR PROJECT

The SOVARR 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. Moreover, if different tools were used in the same experiment, the outputs typically need conversion to some common format, and for reproducibility, this glue code needs to evolve with the

changes of the tools themselves. Similar problems arise with the release of data sets like the Million Song Dataset [1] or the SALAMI dataset [5], 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.

3. LATE-BREAKING SESSION

The late-breaking session on Shared Open Vocabularies started with a discussion on the feasibility of creating shared ontologies within research communities such as the MIR community. It was commonly agreed that the creation of a single shared vocabulary is a difficult challenge and a more feasible alternative is exploring the possibility of several shared modules that are more domain or task specific, that is, vocabularies that serve a specific task in music information research such as genre classification, or more fine tuned to a specific tool or a part of the community with specific research practices. 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 such 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.

As a possible development path, Kevin Page suggested the examination of particular MIR tasks or workflows, such as the ones run by the Music Information Retrieval Evaluation eXchange (MIREX). This could lead to creating vocabularies and ontological models that support particular workflows. There was also general agreement on the importance of demonstrators employing shared open vocabularies in real-world scenarios as they greatly facilitate the adoption of this technology. An example of such a demonstrator built during the Networked Environment for Music Analysis (NEMA) project was discussed, which demonstrates the utility of ontologies, such as the Music Ontology, and Linked Open Data in a typical MIR workflow involving data collection and publishing as well as signal processing [4]. Sonic Annotator Web Application (SAWA) [2] is another demonstrator that was built to help researchers to learn about an existing ontology for audio features. Rudolf Mayer from the Institute of Software Technology and Interactive Systems (ISIS) at the Vienna University of Technology suggested to look at advancements in the Digital Preservation community. This field was originally concerned with the preservation of static digital objects such as multimedia documents to make them resilient to the rapid changes in storage and information access technologies. In this context, an emerging topic is the preservation of complex dynamic digital objects such as workflows in scientific research. In their recent paper [3], Mayer and Rauber argue that previously proposed benchmarking environments, such as MIREX and decentralised frameworks developed in other research communities, do not support the documentation of process execution during experiments. The idea of sharing reusable workflows is an important motivating use case for creating shared vocabularies. Therefore, use cases arising from the preservation of these workflows are indeed important to consider in the SOVARR project.

An enticing possibility emerging from the above requirements is the creation of a web-based service where members of the research community could register their audio feature extraction techniques. The registered methods would receive a unique identifier, and they could be linked with computational algorithms, publications and other relevant metadata. Such a service could also be the basis for resolving terminological differences. For instance, equivalence relations could be established between feature representations with different names as long as they follow the same computational steps. However, it remains difficult to decide who should have authority over creating such links.

4. CONCLUSIONS

The discussions at the Shared Open Vocabularies session clearly demonstrated that there is general consensus on the need for sustainable and interoperable workflows in research communities. However, to bring these ideas into reality, it is essential to collaborate with the entire community making it aware of the tremendous opportunities lying ahead. In this context, both the Semantic Media and the SOVARR projects described in this paper present major community-focused efforts towards achieving interoperability by combining information retrieval workflows with semantic media technologies.

5. REFERENCES

[1] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. 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] G. Fazekas, Y. Raimond, K. Jakobson, and M. Sandler. An overview of semantic web activities in the OMRAS2 project. *Journal of New Music Research (JNMR)*, 39(4), 2010.

[3] R. Mayer and A. Rauber. Towards time-resilient MIR processes. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 337–342, Porto, Portugal, 2012.

[4] K. Page, B. Fields, B. Nagel, G. O'Neill, D. De Roure, and T. Crawford. Semantics for music analysis through linked data: How country is my country? *Sixth IEEE e–Science conference (e-Science 2010), Brisbane, Australia*, 2010.

[5] J. B. L. Smith, J. A. Burgoyne, I. Fujinaga, D. De Roure, and J. S. 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.