Towards a Framework for the Discovery of Collections of Live Music Recordings and Artefacts on the Semantic Web

Thomas Wilmering, Florian Thalmann, Mark B. Sandler
Centre for Digital Music
Queen Mary University of London
{t.wilmering, f.thalmann, mark.sandler}@qmul.ac.uk

ABSTRACT

This paper introduces a platform for the representation and discovery of live music recordings and associated artefacts based on a dedicated data model. We demonstrate our technology by implementing a Web-based discovery tool for the Grateful Dead collection of the Internet Archive, a large collection of concert recordings annotated with editorial metadata. We represent this information using a Linked Data model complemented with data aggregated from several additional Web resources discussing and describing these events. These data include descriptions and images of physical artefacts such as tickets, posters and fan photos, as well as other information, e.g. about location and weather. The system uses signal processing techniques for the analysis and alignment of the digital recordings. During the discovery, users can juxtapose and compare different recordings of a given concert, or different performances of a given song by interactively blending between them.

1. INTRODUCTION

Collecting information from numerous independent sources on the World Wide Web can be cumbersome and its retrieval has to be performed manually. Conventional databases, still most commonly used, do not define the meaning of their data which makes it difficult to connect and interpret information about the same entities at different locations on the Web. For instance, the Grateful Dead collection in the Live Music Archive (LMA) of the Internet Archive[1] provides audio recordings of live music recordings annotated with unstructured metadata. Other resources, such as the Grateful Dead Archive (GDA) curated at the University of California Santa Cruz[2], use their own separate data structure for the description of their content. If one attempted to automatically gather information about a particular concert event from both the LMA and the GDA, one would have to find out how that event is referred to in both archives, and then proceed the same way with entities associated with the event.

The research supporting the project described in this paper is motivated by the hypothesis that Linked Data combined with Semantic Audio[3] technologies can produce novel and improved user experiences for the exploration of large live music recording archives. In this work we combine information such as editorial metadata, audio analysis data, musicological descriptions, as well as digital representations of artefacts such as concert tickets, photos and memorabilia, and making it available on the Semantic Web. In particular, we focus on the interactive exploration of the audio material itself, using dynamic and object-based audio techniques implemented using the Web Audio API[4]. We illustrate this by means of an experimental Web application for the exploration of Grateful Dead shows based on the collection of the Internet Archive, motivated by the richness and completeness of online collections and the continuing scholarly interest in the band's history, regarding both their music, as well as their cultural impact [4, 16].

We first give an overview of the content of the Internet Archive Grateful Dead Collection and briefly introduce Semantic Web Technologies. This is followed by a description of the data model we employ in the implementation of our Web application, and an overview of the type of information we incorporate in its database. We then focus on the procedure we devised to automatically align individual segments of different recordings of the same concert, resulting in a reference timeline for a given concert. We discuss the features of the prototypical Web application that facilitates an audiovisual exploration of the collection before concluding and outlining future work.

2. BACKGROUND

2.1 The Internet Archive Grateful Dead Collection

The Live Music Archive (LMA)[5], part of the Internet Archive, is a growing openly available collection of over 100,000 live recordings of concerts, mainly in rock genres. Each recording is accompanied by basic unstructured metadata describing information including dates, venues, set lists and the source of the audio files. The Grateful Dead collection is a separate collection, created in 2004, consisting of

Web Audio Conference WAC-2017, August 21–23, 2017, London, UK.

---

[1] https://archive.org/details/GratefulDead

[2] https://www.gdao.org

[3] Semantic Audio consists in the extraction of meaning from audio material, such as the musical content, speech content, structure, emotion, similarity, or the layout of the auditory scene.

[4] https://www.w3.org/TR/webaudio/

[5] https://archive.org/details/etree

both audience-made and soundboard recordings of Grateful Dead concerts.

A large number of shows is available in multiple versions. At the time of writing the Grateful Dead collection consisted of more than 11,000 items, recorded on over 2,000 dates spanning the years 1965 to 1995. The late 1960s saw a rise in fan-made recordings of Grateful Dead performances. Indeed, the band encouraged the recording of their concerts for non-commercial use, in many cases providing limited dedicated *taper tickets* for their shows. The *tapers* set up their equipment in the audience space, typically consisting of portable, battery-powered equipment including a cassette or DAT recorder, condenser microphones, and microphone pre-amplifiers. Taping and trading of live shows, especially those of the Grateful Dead, evolved into a subculture with its own terminology and etiquette [15]. The Internet Archive Grateful Dead collection contains digital transfers of such recordings.

2.2 Semantic Web Technologies

We store and retrieve the metadata about the live music recordings along with additional information that relates to the music event using Semantic Web technologies.

These include the Resource Description Framework (RDF) – a data model describing statements as triples of the form *subject, predicate, object.* Combined, these form an RDF graph. Identifying the elements of RDF statements by uniform resource identifiers (URI) creates a globally distributed database, the Web of Linked Data.

Ontology languages like the RDF Schema language (RDFS) and the Web Ontology Language (OWL) [1] allow for defining relationships between entities. They describe a shared conceptualisation of a world [12] consisting of individuals, classes and relations, with formal semantics that allow automated reasoning over RDF data expressed using an ontology. By using these technologies our system supports the seamless linking to and from external data sources. The query language SPARQL allows performing queries over the data by specifying patterns of interlinked triples – it is to triple stores what SQL is to tables & databases.

2.3 Previous Work

In previous work a service that publishes the unstructured metadata from the LMA as Linked Data has been created [3]. The CALMA (Computational Analysis of the Live Music Archive) project [17] supplements this performance metadata with automated computational audio analysis results using feature extraction algorithms. The audio features include high level descriptors such as chords and song tempo, as well as lower level features, among them chroma features [2] and MFCCs [6] which have been used for measuring audio similarity [10, 13] and are of particular interest for the work discussed in this paper. The published CALMA data enables the further investigation of the audio collection, for instance, one may want to investigate musicological aspects by processing and comparing high level audio features. Moreover, the audio feature data can assist in the validation of user-generated collection metadata [17].

In [21] we presented a system that aligns and clusters individual songs from different recordings of a given concert based on various audio characteristics and editorial metadata. It creates an immersive virtual space that can be imported into a multichannel Web or mobile application allowing listeners to navigate the space using interface controls or mobile device sensors.

Related systems have been presented for classical music performances, e.g. [11, 20]. However, these systems are not based on linked data and focus on musical content and include scores rather than concert artefacts.

3. DATA REPRESENTATION

The data model underpinning our Web application reuses various Semantic Web ontologies, such as the Music, Timeline and Event Ontologies [19], the Studio Ontology [9], PROV-O[6], and the CIDOC Conceptual Reference Model (CIDOC CRM) [8]. Figure 1 shows an excerpt of the data model we developed for the application. A Grateful Dead show is represented as an *event* as defined in the Event Ontology. The music performance forms a sub-event, defined as a *performance* in the Music Ontology . The performance is linked to concerts described in the Etree Linked Data service[7]. This allows the easy integration of the Internet Archive metadata and CALMA audio feature data[8] into our System. Other resources include meteorological databases to retrieve historical weather information, and geographical databases to include data about the location a live music event took place.

3.1 Audio Source and Lineage

Our data model supports the detailed description of lineage information using the Studio Ontology, which allows us to model the exact signal flow from recording with a specific microphone model onto a specific recording device to the playback device used for transferring the recording to the digital domain. A given recording may have been made using two *Sennheiser 421* microphones and a *Sony TC-D5M* tape recorder. The Studio Ontology provides concepts for the description of audio equipment[9], which allows us to identify recordings based on recording techniques. We are also able to describe the location of the microphones in relation to the sound sources.

3.2 Provenance

We integrate provenance data, i.e. information about agents that played a role in the creation or modification of a resource with PROV-O, a widely used ontology for expressing provenance information. Agents may be people, such as those who recorded a performance, transferred an analogue recording to the digital domain, created photos or artwork connected to a live show, or software agents, such as audio feature extraction tools (the latter is further described in the context of layered digital libraries and workflow provenance in [17]).

3.3 The Reference Timeline

We introduce the concept of a *reference timeline* in our data structure in order to model the alignment of different recordings of the same performance. Each recording in the collection is provided as separated files reflecting the playlist, however, segment boundaries for the audio files in different

---

[6] https://www.w3.org/TR/prov-o/

[7] http://etree.linkedmusic.org/

[8] http://calma.linkedmusic.org/

[9] for instance, microphones have characteristics like pick-up pattern and frequency response

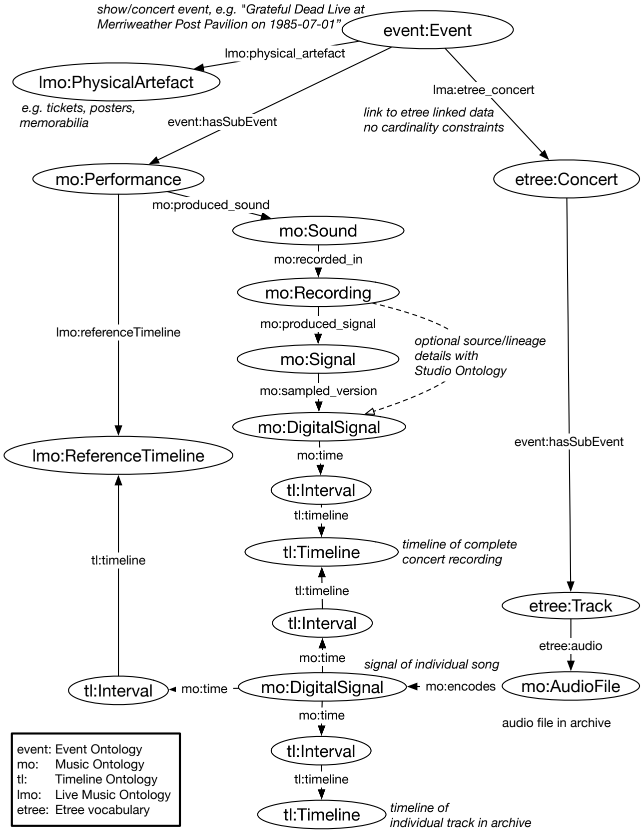


Figure 1: Partial data model for the description of live music recordings in a collection reusing various Semantic Web ontologies.

versions are often inconsistent. This is due to the fact that the start and end times of songs in a live concert can be ambiguous. Moreover, the time between songs, often filled with announcements or instrument tuning, is handled differently across the versions. Some segmentations use separate audio files for these sections, while in other versions the sections are appended or prepended to the previous or following track respectively. The reference timeline allows for the identification of specific points in time of a performance in the segmented recordings available in the collection, facilitating a convenient navigation and comparison of these recordings Figure 2 shows individual tracks from different recordings placed on the reference timeline. In Section 4.1 we discuss the procedure we use to automatically construct the reference timeline by analysing the recordings.

3.4 Physical Artefacts

Artefacts related to a live music event, such as tickets, posters, and other memorabilia, play a significant role in the live music experience and their documentation [5, 14]. We identify photos and images of such physical artefacts from various Web sources, and link them to live music events with the *physical_artefact* property, a sub-property of the *factor* property of the Event ontology. By querying the triple store, digital images depicting these artefacts are retrieved and displayed within our application. Additionally, we map concepts to CIDOC CRM, enabling the detailed description of physical objects and images, and the mediation between different sources of cultural heritage information, such as that published by museums, libraries and archives [8].

4. INFERRING FROM AUDIO FEATURE DATA

At the current stage of the project, we implemented various algorithms that use the feature data gathered in the context of CALMA (Section 2.3) to infer higher-level information about the musical content of the audio. For this, we chose two types of subsets of the musical material to be analysed: all recordings of each concert and all performances of each song.

4.1 Pairwise Alignment

There are up to 22 different versions of recordings for each show, with an average of 5.2 recordings [21]. Due to the nature of the recording equipment, e.g. tape recorders, these recordings often differ in speed and quality, which poses a particular problem for their alignment. They are typically recorded from different positions in the audience which results in varying amounts and fluctuations of crowd noise and a different mix. Many of the recordings, as mentioned above, are segmented differently into songs or groups of songs due to manual processing within a DAW during digitization. Also, different parts may be missing from each of the recordings due to tape switching at different moments.

All these characteristics had to be considered during the alignment procedures. First, we created a pairwise alignment method that dealt with the problem of varying tape speed. We experimented with different alignment methods and decided to use dynamic time warping in the form of the MATCH Vamp plugin [7][10]. We designed a three-step alignment process for pairs of segments which consists in iterative application of the algorithm. First, we apply MATCH to get a sense of the tape speed variation and tuning difference of the pair, then we resample one of them to obtain matching speed and tuning, and finally reapply the MATCH algorithm to get a final alignment in the form of corresponding timepoints.

4.2 Meta-Alignment

We designed an algorithm based on the pairwise alignment algorithm to align n longer sequences of recorded segments $(s^i_0, \ldots, s^i_{k_i})$ for $0 \leq i \leq n$, which is able to deal with the problem of gaps and differing segmentation. The algorithm iteratively constructs a reference timeline that can be used to represent the temporal relationships between the segments (see Section 3). The algorithm first compares all of the first segments $s^i_{j_i}$ for $j_i = 0$ pairwise and constructs the earliest time point of the reference timeline t_0 by measuring the average offset between the segment beginnings, thereby only considering the segment pairs whose alignment works reasonably well[11]. If one of the segments does not rate well in any alignment combination, we proceed to the next one, i.e. increment j_i for segment i, and so on, until a maximum distance in index is reached (for maximum ef-

---

[10] https://code.soundsoftware.ac.uk/projects/match-vamp

[11] Each alignment is rated by calculating the coefficient of determination for a linear regression (R-squared). We define a minimum threshold above the value must exceed, empirically determined as 0.999.

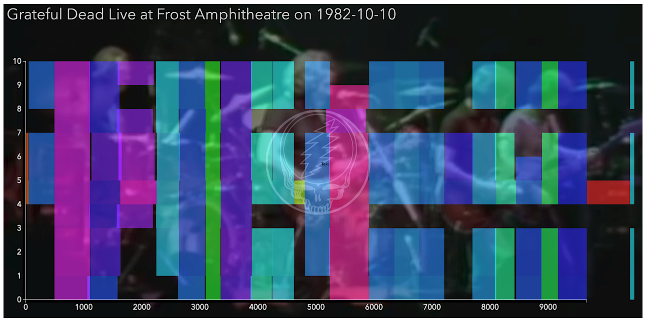


Figure 2: View of different recordings (y-axis) of one concert. Blocks represent segments of the recordings aligned on the constructed reference timeline (x-axis).

ficiency set to 2 or 3). Once the best alignments have been found for the current j_i, we increment all j_i by one, and repeat the procedure. At each iteration we keep track of all optimal alignment pairs and calculate the average offsets between the segments. From this, we infer their position on the reference timeline. Figure 2 shows a visualisation of the result for the meta-alignment of all recordings of a selected Grateful Dead concert.[12]

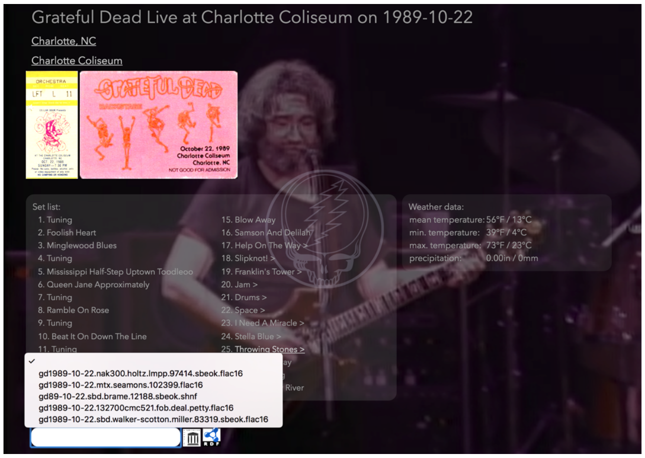
4.3 Spatial Clustering

As mentioned in Section 2.3, we described another inference procedure in [21]. It consists in clustering the different recordings of a song based on the dissimilarity of their local temporal content calculated from the average distance of their chroma and MFCC features. This resulted in a spatial arrangement of the individual channels of the recordings which could then be placed in binaural space and navigated interactively. Using the methods presented in the previous section, we are now able to calculate these clusterings reliably for an entire concert.

5. WEB APPLICATION PROTOTYPE

Figures 2 and 3 show screenshots of the prototypical Web application for the discovery of the Grateful Dead collection. Each concept in the ontology, e.g. a concert, venue, geographical location, or song, has a particular corresponding view with optional more detailed audio-based views. Figure 3, for instance, shows the page for a Grateful Dead show at a certain date. The menu on the lower left lists the available recordings of the performance available in the Internet Archive. Scans of tickets and historical weather information are automatically displayed on the page. Each page has numerous links which directly lead to related groups of concepts. For example, from the venue page one can directly get to all the concerts that have taken place at that location. The pages are created dynamically based on information retrieved by SPARQL queries that are performed in the background.

Figure 2 shows a more detailed view with visualisation that juxtaposes all the recordings of a particular concert. All audio-based views have a playback mode, where different aspects of the audio can be heard and recombined. In the view in Figure 2, for example, one can listen to the concert while smoothly blending between the different recordings. In another concert-based detail view, introduced in [21], one can listen to the immersive clustered version of the recordings and move around interactively.


Grateful Dead Live at Charlotte Coliseum on 1989-10-22

Charlotte, NC

Charlotte Coliseum

ORCHESTRA
LFT L 11

October 22, 1989
Charlotte Coliseum
Charlotte, NC
NOT GOOD FOR ADMISSION

Set list:
1. Tuning
2. Foolish Heart
3. Minglewood Blues
4. Tuning
5. Mississippi Half-Step Uptown Toodleoo
6. Queen Jane Approximately
7. Tuning
8. Ramble On Rose
9. Tuning
10. Beat It On Down The Line
11. Tuning
15. Blow Away
16. Samson And Delilah
17. Help On The Way >
18. Slipknot! >
19. Franklin's Tower >
20. Jam >
21. Drums >
22. Space >
23. I Need A Miracle >
24. Stella Blue >
25. Throwing Stones >

Weather data:
mean temperature: 56°F / 13°C
min. temperature: 39°F / 4°C
max. temperature: 73°F / 23°C
precipitation: 0.00in / 0mm

✓
gd1989-10-22.nak300.holtz.lmpp.97414.sbeok.flac16
gd1989-10-22.mtx.seamons.102399.flac16
gd89-10-22.sbd.brame.12188.sbeok.shnf
gd1989-10-22.132700cmc521.fob.deal.petty.flac16
gd1989-10-22.sbd.walker-scotton.miller.83319.sbeok.flac16

Figure 3: Screenshot of the Web application.

Not shown here is an "expert view" where the user can construct custom queries in order to retrieve audio material based on a combination of characteristics. For example, one could list *all available versions of the song "Dark Star" recorded with microphones with cardioid pick-up pattern at shows that took place in open-air venues between 1970 and 1975.* In future versions, queries may also support audio features from the CALMA database.

6. CONCLUSIONS AND FUTURE WORK

The data model and prototype presented in this paper can easily be extended to include other audio-related information as well as other artefacts and information taken from various sources. Similar query, navigation and playback concepts can be used for other semantic audio relationships. For instance, we are currently working on structural inference algorithms that will help us compare and align different performances of the same song over time, as well as further extending the Music Ontology framework with a dedicated Live Music Ontology.

While our data model is capable of representing multifaceted information in high detail, its use in applications relies on equally detailed accurate metadata, which in many cases is not present. The descriptions of the music collection we base our use case on for instance, comprises of historical audio recordings with mostly incomplete unstructured user-generated metadata. However, the data model and audio processing techniques discussed in this paper can also be used for future projects where metadata of live music recordings is collected at the point of creation.

7. ACKNOWLEDGMENTS

This paper is supported by EPSRC Grant EP/ L019981/1, Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption. Mark B. Sandler acknowledges the support of the Royal Society as a recipient of a Wolfson Research Merit Award.

---

[12] Prätzlich and Müller [18] presented a similar alignment procedure for opera performances. However, instead of constructing a reference timeline, they used a reference segmentation created by musicologists.

8. REFERENCES

[1] G. Antoniou and F. van Harmelen. Web ontology language: OWL. *in S. Staab and R. Studer (eds.), Handbook on Ontologies, International Handbooks on Information Systems, Springer-Verlag Berlin Heidelberg*, pages 91–110, 2009.

[2] M. A. Bartsch and G. H. Wakefield. To catch a chorus: Using chroma-based representations for audio thumbnailing. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 15–18, New Paltz, NY, USA, 2001.

[3] S. Bechhofer, K. Page, and D. De Roure. Hello Cleveland! Linked Data Publication of Live Music Archives. In *Proceedings of WIAMIS, 14th International Workshop on Image and Audio Analysis for Multimedia Interactive Services*, 2013.

[4] M. Benson. *Why the Grateful Dead Matter*. ForeEdge Press, 2016.

[5] K. Burland and S. Pitts, editors. *Coughing and Clapping: Investigating Audience Experience*. Routledge, 2016.

[6] S. Davis and P. Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4):357–366, 1980.

[7] S. Dixon and G. Widmer. Match: A music alignment tool chest. In *ISMIR*, pages 492–497, 2005.

[8] M. Doerr. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine*, 24(3), 2003.

[9] G. Fazekas and M. B. Sandler. The studio ontology framework. *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 2011.

[10] J. T. Foote. Content-based retrieval of music and audio. In C.-C. J. Kuo, S.-F. Chang, and V. N. Gudivada, editors, *Multimedia Storage and Archiving Systems II*, 1997.

[11] M. Gasser, A. Arzt, T. Gadermaier, M. Grachten, and G. Widmer. Classical music on the web-user interfaces and data representations. In *ISMIR*, pages 571–577, 2015.

[12] M. R. Genesereth and N. J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers, San Mateo, CA, USA., 1987.

[13] B. Logan and A. Salomon. A music similarity function based in signal analysis. *IEEE International Conference on Multimedia and Expo (ICMC)*, 2001.

[14] M. McCray Pattacini. Deadheads yesterday and today: An audience study. *Popular Music and Society*, 24(1):1–14, 2000.

[15] N. Meriwether. Documenting the dead. *Online: http://www.dead.net/documenting-the-dead*, 2015.

[16] U. Olsson. *Listening for the Secret: The Grateful Dead and the Politics of Improvisation*, volume 1. Univ of California Press, 2017.

[17] K. R. Page, S. Bechhofer, G. Fazekas, D. M. Weigl, and T. Wilmering. Realising a layered digital library. In *Proceedings of the 17th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'17)*, 2017.

[18] T. Prätzlich and M. Müller. Frame-level audio segmentation for abridged musical works. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Taipei, Taiwan, 2014.

[19] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The music ontology. *Proceedings of the International Conference on Music Information Retrieval*, 2007.

[20] D. Röwenstrunk, T. Prätzlich, T. Betzwieser, M. Müller, G. Szwillus, and J. Veit. Das gesamtkunstwerk oper aus datensicht. *Datenbank-Spektrum*, 15(1):65–72, 2015.

[21] T. Wilmering, F. Thalmann, and M. B. Sandler. Grateful live: Mixing multiple recordings of a Dead performance into an immersive experience. In *Proceedings of the Audio Engineering Society Convention 141*, 2016.