

FAIR streams of MIDI Linked Data

Albert Meroño-Peñuela, Computer Science Department, Vrije Universiteit Amsterdam

@albertmeronyo

Category: short paper

Abstract. In order to address the urgent need of improving the infrastructure for reusing scholarly data, the FAIR guiding principles for scientific data management and stewardship [1] have been proposed to concisely measure the Findability, Accessibility, Interoperability and Reusability of scientific data. This has fueled research investigating the extent to which these FAIR principles are satisfied in domain-specific data cultures. A recent example of such a domain is music and musicology: in repositories and digital libraries, digitizations of scores and lyrics (77%) in PDF and/or JPG (40%) are the most common data artifacts; and the more the scale of the repository increases, the less structured formats for representing symbolic notation seem to be used [2]. The lack of machine-readability of symbolic musical data hampers the adherence to the FAIR principles. Given this situation for music that is considered worth of long-term archival, it is only expected that representations of **live performances**, of which data is generally harder to find, access, interoperate and reuse, suffer the problem to the same or a greater extent.

In this paper, and based on previous research results, we address the problem of making symbolic MIDI music representations FAIR by representing music in the early stage of its inception, the *live performance*, as de-referenceable RDF triples. As a result, our approach can generate FAIR streams of MIDI Linked Data with low latency that enable users to find, access, interoperate and reuse MIDI music in an unprecedented manner. Our approach uses a state-of-the-art generative audio matching algorithm [4] to generate a stream of MIDI events from analogic musical instruments; and the *midi2rdf* mapping algorithm [3] to map these MIDI events into RDF triples.

We argue that these MIDI data streams of RDF triples are a fundamental advance for making live music performance data FAIR. First, music performances become Findable thanks to unique and persistent identifiers (URIs) for all its notes, instruments, and events. Second, they become Accessible since these URI identifiers can be dereferenced through the HTTP protocol in order to obtain more data. Third, they become Interoperable through the use of the MIDI ontology vocabulary [6], allowing communication exchange with other datasets. And last, they become Reusable through open licensing and the addition of relevant provenance metadata. Our expectation is for these FAIR streams of MIDI Linked Data to become central in the task of content-to-metadata linkage [5,6] and to enable their indexing and communication through Linked Data Notifications [7].

References

1. Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016).

2. Marilena Daquino, Enrico Daga, Mathieu D'Aquin, Aldo Gangemi, Simon Holland, Robin Laney, Albert Meroño-Peñuela, Paul Mulholland. "Characterizing the Landscape of Musical Data on the Web: State of the art and challenges". In: Proceedings of the 2nd Workshop on Humanities in the SEmantic web (WHiSe 2017). ISWC 2017, October 22nd, Vienna, Austria (2017).
3. Albert Meroño-Peñuela, Rinke Hoekstra. "The Song Remains the Same: Lossless Conversion and Streaming of MIDI to RDF and Back". In: 13th Extended Semantic Web Conference (ESWC 2016), posters and demos track. May 29th — June 2nd, Heraklion, Crete, Greece (2016).
4. Kristensen, O.: Generative audio matching game system (Mar 3 2011), <https://www.google.com/patents/WO2010142297A3?cl=en>, wO Patent App. PCT/DK2010/050,132
5. Albert Meroño-Peñuela, Reinier de Valk, Enrico Daga, Marilena Daquino, Anna Kent-Muller. "The Semantic Web MIDI Tape: An Interface for Interlinking MIDI and Context Metadata". In: Workshop on Semantic Applications for Audio and Music, ISWC 2018. 9th October 2018, Monterey, California, USA (2018)
6. Albert Meroño-Peñuela, Rinke Hoekstra, Aldo Gangemi, Peter Bloem, Reinier de Valk, Bas Stringer, Berit Janssen, Victor de Boer, Alo Allik, Stefan Schlobach, Kevin Page. "The MIDI Linked Data Cloud". In: The Semantic Web – ISWC 2017, 16th International Semantic Web Conference. Lecture Notes in Computer Science, vol 10587, pp. 156-164 (2017)
7. Capadisli, Sarven, Amy Guy, Christoph Lange, Sören Auer, Andrei Sambra, and Tim Berners-Lee. "Linked data notifications: a resource-centric communication protocol." In European Semantic Web Conference, pp. 537-553. Springer, Cham, 2017.