The Oldest Song Score in the Newest Notation: The Hurrian Hymn to Nikkal as Linked Data

Introduction

The Web can constitute a natural medium for the publication and discovery of historical evidence, their related resources, and descriptive metadata. Increasingly, but steadily, more and more historical artifacts make their way from archaeological sites to musea, libraries, digital archives[1], and also the Web of data [12], besides their traditional proliferation to the wider public through books [1]. The discovery and retrieval of historical digital objects and their descriptive metadata on the open data space of the Web are challenging tasks [12].

For example, *"What is the oldest music score known?"* is a relevant question of interest to art historians, music historians, musicologists, and the wider public. Web search can lead users to at least a partial answer to this question; as it turns out, the oldest song known left to present in written form is a Sumerian Hymn written 3,400 years ago[2]. Part of the Hurrian songs, the Hurrian Hymn to Nikkal (also known as h.6) is the oldest substantially complete work of notated music in the world, inscribed in cuneiform on clay tablets and excavated from the ancient city of Ugarit (today northern Syria) and dated 1400 BC [2]. The work of a number of historians has allowed for these inscriptions to be transcribed in modern Western notation [3], ultimately leading to the possibility of being played in a modern lyre,[3] which resembles ancient Sumerian instruments.[4]

The provenance trail of this unprecedented discovery of musical culture is complex, and this complexity manifests on the Web in the form of variety and heterogeneity. Necessary pieces of knowledge to reconstruct this breakthrough are scarce, hard to find, and, most importantly, spread through a number of heterogeneous and semantically incompatible information representation formats: HTML hypertext, PDF documents, scanned images, digital score transcriptions, MIDI files, MP3 files, etc.[5] Therefore, the complexity of all human available Web knowledge on h.6 comes in the form of multimodality.

In the broader context of digital data access and processing, we must therefore ask: how can these relevant, multimodal pieces of knowledge be queried together for further computation in a consistent and reproducible way? In this paper, and following established practice [12], we propose to use the Resource Description Framework (RDF), the Linked Data paradigm, and the SPARQL query language [4] to answer fundamental questions in music history about the Hurrian Hymn to Nikkal. In order to do so, we follow a three-step approach in which, first, we convert MIDI representations of the hymn into RDF, effectively transcribing the notation of the oldest known music written score into a modern, well understood, and machine processable knowledge representation language; second, we enrich this RDF representation with all provenance metadata facts we could collect from the Web about the hymn, its discovery, and its interpretation; and, third, we store data retrieving

---

[1] https://easy.dans.knaw.nl/ui/home
[2] http://www.openculture.com/2014/07/the-oldest-song-in-the-world.html
[3] http://www.ancientlyre.com/
[4] https://www.youtube.com/watch?time_continue=5&v=9c-hmFN610g
[5] http://www.openculture.com/2014/07/the-oldest-song-in-the-world.html

SPARQL queries to provide unique API links that reproduce the retrieval of relevant h.6 data without knowledge of these technologies. We use the results of this methodology to study the relevance and idiosyncrasy of the hymn when we compare its notation and metadata to a large dataset of modern MIDI music of 10 billion RDF facts.

Background

The Hurrian songs are a collection of music inscribed in cuneiform on clay tablets. These tables were originally excavated in the 1950s from the ancient city of Ugarit, a headland in northern Syria, and date to approximately 1400 BC. The Hurrian hymn to Nikkal (also known as the Hurrian cult hymn or A Zaluzi to the Gods, or simply h.6) is encoded in one of these clay tablets, and it was first transcribed into modern Western notation in 1972 [2] (see Figure 1). After this, a variety of alternative interpretations have been suggested [3]. The proliferation of various plausible interpretations and theories on the correspondence of h.6's content and modern music notation has led to different modern recordings.[6] Many of these are scarcely available [5], while some others are openly available on the Web.[7] The use of the lyre is generally accepted as a proxy to recreate the timbre of original Sumerian instruments. Other transcriptions use the popular synthesizer language MIDI.
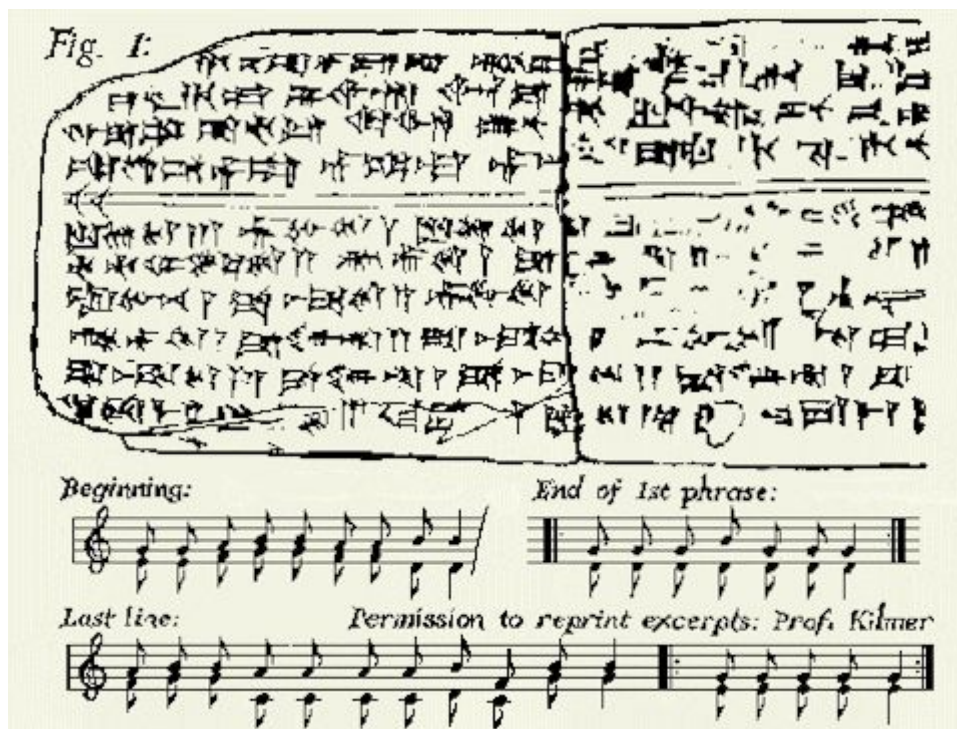

Figure 1. Excerpt of the Hurrian hymn to Nikkal and its transcription into modern Western music notation [2].

The Musical Instrument Digital Interface (MIDI) [6] standard allows electronic musical devices to communicate by exchanging messages that can be interpreted as music notation. MIDI encodes so-called "events" into typically 3-byte messages that describe some event

---

[6] We must note that the goal of this paper is not to contest these interpretations; but to make their arguments, contents, resources and reasoning workflows available in Web open, standard and machine-processable formats.
[7] https://www.youtube.com/watch?time_continue=5&v=9c-hmFN610g

relevant for the production of musical sound. For example, the action of pressing the middle C key in the piano quickly can be expressed with the MIDI message <144, 60, 100> (144 is 90 in hexadecimal; 9 stands for the type of event, "note on" or "start sounding a note"; 0 for the first MIDI channel; 60 is the 60th key in the piano counting from left to right; and 100 is a measure of how hard that key is hit).

The midi2rdf algorithm [7] represents information originally encoded as MIDI in the open, standard and machine-processable knowledge representation language of the Web: the Resource Description Framework (RDF) [8]. RDF expresses knowledge as subject-predicate-object sentences (or "triples"). RDF triples use URIs identifiers (the same identifiers used to uniquely identify HTML pages on the Web) to indicate these subjects, predicates and objects. For example, the fact that "Tim Berners-Lee is a person" can be expressed in RDF with the triple <https://www.w3.org/People/Berners-Lee/> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person>. Therefore, the midi2rdf algorithm can be used to transform any MIDI file into a collection of such RDF triples. This has been done on a large collection of 500K MIDI files gathered from the Web, leading to the creation of the MIDI Linked Data Cloud, the largest dataset of machine-readable music notation, containing more than 10B RDF triples about MIDI songs and all their events and notes [9]. More recent work proposes methods that leverage musicians' performances, MIDI similarity algorithms, and entity recognition techniques to establish links between music notation content (e.g., the MIDI notes of "Hey Jude" by the Beatles) and their descriptive Web metadata (e.g. the Wikipedia page of that song[8]) [10].

Approach and Findings

Here, we propose to follow a three-step method in order to represent historical symbolic music notation content and descriptive metadata of the MIDIs of the Hurrian Hymn to Nikkal as Linked Data through RDF, and incorporate those into the MIDI Linked Data Cloud:

1. Symbolic music notation content. As a first step, we encode the MIDI files of the Hurrian Hymn to Nikkal in RDF, using the midi2rdf algorithm. This produces various graphs of RDF data encoding the different musical events of the digital notation.

2. Descriptive metadata. In this step, we add additional contextual provenance and metadata triples to the RDF graphs produced in the previous step. These are harder to integrate in the historical case than in modern music, since generic approaches linking only large datasets is ineffective in this case. To address this, we investigate the resources and workflows around the creation of modern music notations of the h.6, and compile all these using newly minted URIs, shared vocabularies, and reusable modeling practices.

3. Reproducible queries. Since the output of the previous steps is RDF data that can only be queried through the (complex) SPARQL query language, we compile a list of SPARQL queries encoding relevant competency questions over these new RDF graphs, interrogating them using combinations of musical knowledge (derived from the MIDIs) and contextual provenance (derived from the links compiled in step 2). By doing so next to automatic API generation tools [11], we generate stable and

---

[8] https://en.wikipedia.org/wiki/Hey_Jude

reproducible links to execute these queries, retrieving predictable results without the need of knowledge in RDF or SPARQL.

Our findings reveal both shortcomings and advantages to this approach. The first limitation is the manual, unprincipled, and scarcely understood nature of collecting Web-based relevant provenance and contextual resources, links and workflows over these music-historical objects. Secondly, this process is prone to be biased towards the selection of resources that are only available on the Web, ignoring those published elsewhere; this aims at the general problem of reachability of offline archives and repositories. These shortcomings are balanced by a number of advantages. First, the end result supports a thorough documentation process that generates a semantic RDF graph of historically relevant and connected Web resources, posing the added value of machine-readability over more traditional text-based documentations. Second, the result supports a more principled retrieval mechanism, based on mixing musical (through the h.6 notes represented in RDF) and historical (through the supporting metadata RDF graph) knowledge under the same querying paradigm (reproducible SPARQL links). Third, the availability of both MIDI and their history enables a more exploratory approach, in which besides providing query results, the serendipitous discovery of information is incentivated through similarity links. Ultimately, we hope to enable comparative studies through graph metrics pointing differences between historical and contemporary music.

References

1. Anne Kilmer and Sam Mirelman. "Mesopotamia". In: Grove Music Online, Oxford Music Online. The New Grove Dictionary of Music and Musicians, second edition, edited by Stanley Sadie and John Tyrrell https://doi.org/10.1093/gmo/9781561592630.article.18485
2. Anne Draffkorn Kilmer, "The Discovery of an Ancient Mesopotamian Theory of Music", Proceedings of the American Philosophical Society Vol. 115, No. 2 (1971), pp. 131-149
3. Marcelle Duchesne-Guillemin. "A Hurrian musical score from Ugarit: the discovery of mesopotamian music". Sources from the Ancient Near East, volume 2, fascicle 2 (1984)
4. Steve Harris, Andy Seaborne, and Eric Prudhommeaux. "SPARQL 1.1 query language". W3C recommendation, 21(10) (2013).
5. Anne Kilmer, Richard Crocker. "Sounds from silence: Recent discoveries in ancient Near Eastern music". Bit Enki publications (1976)
6. The MIDI Manufacturers Association: "The Complete MIDI 1.0 Detailed Specification". Tech. rep., The MIDI Manufacturers Association, Los Angeles, CA (1996- 2014), https://www.midi.org/specifications/item/the-midi-1-0-specification
7. 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)
8. Richard Cyganiak, David Wood, Markus Lanthaler. "RDF 1.1 Concepts and Abstract Syntax". W3C Recommendation 25 February 2014 (2014) https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/

9. 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)
10. 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)
11. Albert Meroño-Peñuela, Rinke Hoekstra. "Automatic Query-centric API for Routine Access to Linked Data". In: The Semantic Web – ISWC 2017, 16th International Semantic Web Conference. Lecture Notes in Computer Science, vol 10587, pp. 334-339 (2017)
12. Albert Meroño-Peñuela, Ashkan Ashkpour, Marieke van Erp, Kees Mandemakers, Leen Breure, Andrea Scharnhorst, Stefan Schlobach, Frank van Harmelen. "Semantic Technologies for Historical Research: A Survey". Semantic Web — Interoperability, Usability, Applicability, 6(6), pp. 539–564. IOS Press (2015)