

On the Adoption of Standard Encoding Formats to Ensure Interoperability of Music Digital Archives: The IEEE 1599 Format

A. Baratè, L. A. Ludovico, F. Simonetta

Department of Computer Science
University of Milan

{adriano.barate,luca.ludovico,federico.simonetta}@unimi.it

D. A. Mauro

Department of Computer & Information Technology
Marshall University
maurod@marshall.edu

ABSTRACT

With this paper, we want to stimulate the discussion about technologies for inter-operation between various music datasets and collections. Among the many standards for music representation, IEEE 1599 is the only one which was born with the exact purpose of representing the heterogeneous structures of music documents, granting full synchronization of all the different aspects of music (audio recordings, sheet music images, symbolic representations, musicological analysis, etc). We propose the adoption of IEEE 1599 as an interoperability framework between different collections for advanced music experience, musicological applications, and Music Information Retrieval (MIR). In the years to come, the format will undergo a review process aimed at providing an updated/improved version. It is now the perfect time, for all the stakeholders, to come together and discuss how the format can evolve to better support their requirements, enhancing its descriptive strength and available tools. Moreover, this standard can be profitably applied to any field that requires multi-layer and synchronized descriptions.

CCS CONCEPTS

• **Applied computing** → **Sound and music computing**; • **Information systems** → *Digital libraries and archives*; • **General and reference** → *Computing standards, RFCs and guidelines*.

KEYWORDS

IEEE 1599, XML, multi-layer encoding, digital archives, interoperability

ACM Reference Format:

A. Baratè, L. A. Ludovico, F. Simonetta and D. A. Mauro. 2019. On the Adoption of Standard Encoding Formats to Ensure Interoperability of Music Digital Archives: The IEEE 1599 Format. In *6th International Conference on Digital Libraries for Musicology (DLfM '19)*, November 9, 2019, The Hague, Netherlands. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3358664.3358665>

1 INTRODUCTION

Providing a comprehensive description of music information implies multiple points of view. Catalogue metadata, scores, audio

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DLfM '19, November 9, 2019, The Hague, Netherlands

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7239-8/19/11...\$15.00

<https://doi.org/10.1145/3358664.3358665>

tracks, computer-based formats are only a few examples of the heterogeneity to be managed in order to describe a single music piece in all of its aspects. On one side, such a complexity – if properly managed – can pave the way to a number of innovative and advanced applications, ranging from a more complete and satisfactory music experience to the possibility to conduct multi-layer content analysis. On the other side, even if digitization campaigns have originated a great number of music-related digital objects, the adoption of non-interoperable encoding formats and the geographical distribution across multiple institutions (libraries, archives, repositories, etc.) have often limited the possibility to enjoy heterogeneous information as a whole.

The research question we want to address in this work is based on these considerations. We believe that an effort in the standardization of music representation deriving from multiple and potentially heterogeneous sources can be important for many categories of users, from music enthusiasts to musicians and musicologists, also including the Music Information Retrieval (MIR) community.

In this sense, the IEEE 1599 format represents the perfect match, as it provides the ability to collect and represent in a synchronized way various kinds of information related to a single music piece within a multi-layer environment. Since the IEEE 1599 is under revision, this is the right moment to investigate possible changes concerning interoperability.

The remainder of the paper is structured as follows: Section 2 describes the key features of the IEEE 1599 standard, both in its current form and in its expected evolution; Section 3 discusses the applicability of this format to different categories of musical assets' stakeholders; Section 4 proposes an architecture to federate distributed information sources; finally, Section 5 discusses the most relevant advantages deriving by such an effort, draws the conclusions and sheds some light on future developments.

2 KEY FEATURES OF THE STANDARD

IEEE 1599 is a language for a comprehensive description of music, standardized by IEEE Standards Association in 2008. IEEE 1599 provides a meta-representation of music information within a multilayered environment, which achieves integration among the general, structural, notational, computer-driven performance, and audio layers [2]. Music information encoding often adopts several distinct reference formats for audio (e.g., CD-DA, DVD-A, FLAC, MP3, AAC), for computer-driven performance (e.g., MIDI, MPEG, SASL/SAOL), for music scores (score editors' proprietary formats, MEI, MusicXML). Some of these are formal standards, others represent *de-facto* practices. Each of these deals with musical information only in a restricted sector, and not addressing all its aspects simultaneously. Conversely, in our opinion there is a strong need for the

integration of the various layers musical information is made of (audio, performance, music notation, musical forms, metadata), in order to provide access to all these layers interactively and as an integrated whole. This would enable, for instance, the navigation of score notation while listening to the corresponding audio (score following), the real-time comparison of different graphical representation as well as audio performances, and the interaction with musical contents within a multimodal environment. IEEE 1599 integrates music representation with already defined and commonly accepted standards and formats.

Another key feature of the format is the possibility to support multiple digital objects for each layer, and to achieve synchronization among instances, both belonging to the same layer and across layers, based on time and space dimensions.

The format has been conceived not only to act as an aggregator for heterogeneous music-related contents, but also as an interchange format among different applications. In this sense, the standard addresses any kind of software dealing with music information, e.g. digital score editors, optical music recognition (OMR) systems, web and mobile apps, musical databases and archives, and performance, composition and musicology-oriented applications. IEEE 1599 paves the way for novel applications for music enjoyment, publishing and research, such as innovative multimedia products [7], music-oriented educational platforms [8], and software tools for the maintenance and exploitation of cultural heritage [6].

2.1 The 2008 Standard

IEEE 1599 encodes all music-related information in XML, which is a hierarchical, extensible, portable, and machine and human-readable language. This choice makes it similar to other standards originally conceived to describe and interchange musical notation, such as MusicXML and the Music Encoding Initiative. The main goal of IEEE 1599, however, is different: supporting and synchronizing multi-layered music information, as defined in [13], [17] and [20].

The milestones that brought to the standardization of IEEE 1599 are the establishment of the *IEEE Computer Society Task Force on Computer Generated Music* (1992), the constitution of the *IEEE Technical Committee on Computer Generated Music* (1994), the approval by IEEE Standard Association of the *Recommended practice for the Definition of a Commonly Acceptable Musical Application Using the XML Language* by the IEEE Standard Association, and, consequently, the creation of the *IEEE Standards Association Working Group on Music Application of XML* (2001). The balloting phase ended in 2008, thus making IEEE 1599 an internationally recognized standard.

During the following 10 years, a number of applications have been developed, both to make the production of materials quicker and easier, and to show the potential of the format when applied to multimedia fruition, formal music education, gamification and edutainment, computational musicology, promotion of intangible cultural heritage, etc. Besides, about 50 scientific papers and a book [3] dealing with IEEE 1599 and its applications have been published. On the other hand, some critical issues have emerged as well, as remarked in [4].

In conclusion, if on one side the interest of the scientific community towards the multi-layer approach has been and is still high,

the penetration of the format in industry has been unsatisfactory so far.

2.2 Expected Evolutions

In March 2018, 10 years after the standardization of the first version, IEEE started a new recommended practice to update the standard. The reasons for this new initiative are multiple: first, technology has evolved and other similar initiatives have appeared; moreover, during the last years, both requests for improvement in the definition of the standard and new needs emerged; finally, the research group who defined the format was planning to extend the representation possibilities offered by the original standard [5].

The new recommended practice strives to generalize the organization of the document into layers, letting the user define custom layers too. In this way, IEEE 1599 could provide support to representation domains that are currently unpredictable, overcoming the original rigid 6-layer structure. Other change requests are more focused on specific aspects, nevertheless their impact could be crucial to the affirmation of the standard in an interoperability context. For instance, a future generalization of the Logic layer should natively foster the integration of currently-supported notational formats with new ones. Besides, external media materials linked from the XML document could be organized in a better way, maybe enclosed in a suitable file format to deliver both the XML and the attached binary objects. Another important advancement is the expected integration of Digital Rights Management representation within the IEEE 1599 multilayer asset. It is worth remembering that the format does not aim at substituting the various format or standards already available but rather at providing a meaningful integration.

These possibilities have been investigated and proposed to the scientific community during the 1st International Workshop on Multilayer Music Representation and Processing (MMRP19), held in Milano, Italy on 24-25 January 2019 [10]. On that occasion, the steering committee has organized 3 panels dealing with IEEE 1599, focusing on: (1) history, technical notes, and demos, (2) the goals of the new PAR WG1599, and (3) the structure of the Working Group and a draft of workplan. The discussion, which involved a worldwide audience of experts and scholars, resulted in the constitution of 5 sub-working groups, dealing with (1) Descriptive framework extensions, (2) Automatic recognition, (3) Intellectual property and Digital Right Management, (4) Platform improvements, and (5) Demo deliverables, respectively.

3 APPLICABILITY TO DIGITAL LIBRARIES, REPOSITORIES AND DATASETS

The aim of this paper is to investigate the potential offered by IEEE 1599 to the stakeholders of musical heritage, and, specifically: (1) *music digital libraries*, (2) *archives*, and (3) *dataset repositories*.

Analyzing music and music-related objects hosted by these categories of institutions, a great heterogeneity emerges. In the following, we will treat such a multifaceted subject by applying the multilayer approach mentioned above. The goal is to show how the extensive adoption of IEEE 1599 both as an encoding format and as an information interchange standard could federate multiple geographically-distributed sources in order to provide users with a comprehensive experience of music.

Examples of *music libraries* of our interest include:

- (1) Traditional libraries with a digital catalogue – This is the case of the *Biblioteca Marciana* in Venice,¹ that holds manuscript scores by Antonio Vivaldi, but publishes online catalogue information only;²
- (2) Libraries sharing (also) digital objects – An example is the “Music Classics” section of the *Biblioteca Europea di Informazione e Cultura (BEIC)*, where both catalogue information and musical excerpts are publicly available;³
- (3) Thematic on-line repositories – A relevant example focusing on free public-domain sheet music is the *International Music Score Library Project (IMSLP) – Petrucci Music Library*.⁴

According to the 6-layer layout proposed in IEEE 1599, all the mentioned cases provide general information, i.e. metadata, while Case 2 involves also the audio layer, and Case 3 the notational one.

Digitized *music archives* are another relevant source of heterogeneous information. For instance, *La Scala* of Milan has an integrated management system for its heritage: *La Scala DAM* (Digital Asset Management), that represents the digital archive of all the available material, from the second decade of the '900 to the present, stored in various archives, warehouses and safes. The theater's activity is documented in 5000 tapes containing operas, ballets and symphonic concerts from 1950 onwards (circa 10,000 hours of music), in 17,000 posters, and in more than 1,000,000 photographs taken from the stage, rehearsals and back-stage. Besides, *La Scala's* archives comprises about 24,000 sketches and costume designs, 45,000 costumes completed by 60,000 accessories (jewels, clothing, footwear, wigs and hats), 80,000 props. A rich selection of these materials is freely available at *ArchivioLaScala*, the web page of the DAM.⁵ In this case, a single multimedia database can provide information for multiple layers (general, notational, audio), and also enrich the description of a music piece with additional digital objects (e.g., on-stage photos, playbills, etc.).

Concerning the third typical case of music stakeholders, many *datasets* exist for MIR tasks, but they often lack the ability of inter-operation. For instance, [16] contains source-separated and mixed audio and video tracks, MIDI scores and frame-level transcriptions; in [12] a dataset containing audio recordings, music scores and sheet music images is used; another interesting multimodal dataset containing time-aligned notes, audio and lyrics is presented in [18]; audio recordings, notes and expressive markings were recently collected in [15]. To date, each of these datasets used its own format for representing the synchronization of music along the various modalities. The researcher who wants to use more than one dataset needs to implement multiple reading utilities. Moreover, these *ad-hoc* music formats are not able to deal with the future developments of music technologies, which continuously demand for new kinds of music information. For instance, the same music piece is often used in multiple libraries, but it is hard to merge the annotations coming from the various archives because digitization/labeling efforts were conducted without reference to standard annotations for synchronization.

¹<https://marciana.venezia.sbn.it/>

²<https://polovea.sebina.it/SebinaOpac/query/antonio%20vivaldi>

³<https://www.beic.it/it/articoli/classici-della-musica>

⁴<https://imslp.org/>

⁵<http://www.archiviolascala.org>

4 THE PROPOSED ARCHITECTURE

An in-depth explanation of the mechanisms adopted by IEEE 1599 to support multiple descriptions of music events and keep them mutually synchronized goes beyond the scope of this paper. It is just worth mentioning that each IEEE 1599 document, in a sub-part of the Logic layer, has a list of identifiers for music events, called the *spine*. While the logical descriptions of these events are expressed in XML format and contained inside the document itself, multimedia descriptions (e.g., images and audio tracks) are demanded to external digital objects encoded in usual formats (e.g., TIFF and MP3 files). In other words, the IEEE 1599 document does not embed multimedia, but contains in the suitable layers the information required to retrieve the occurrences of music events inside external files. For instance, the Audio layer links a number of audio performances of the piece, and the position of music events within each of these instances is expressed within the XML document as a time offset from the beginning of the track. Similarly, the Notational layer links external graphical files containing notation, and the music symbols corresponding to spine events are identified through bounding boxes expressed in pixels in the XML code.

This structure where synchronization information is internal to the document and multimedia information is external presents a number of advantages due to the adoption of commonly accepted formats for digital objects: supporting a huge number of already-available media types and formats, relying on their characteristics to encode information efficiently and effectively, avoiding the need to re-encode information into the verbose XML language, etc.

In the context of our research, however, the main advantage is the possibility to link geographically distributed files with no change to the standard. This approach opens the way to a number of innovative applications based on information sharing and integration. The possibilities emerging from the establishment of a network of relationships both internal to an IEEE 1599 document and external, i.e. involving other information sources, has been already addressed in [9] in the context of the Semantic Web.

In the present work we propose a different vision, involving a number of musical assets' stakeholders sharing information in IEEE 1599 format. A possible architecture of the environment is shown in Figure 1. According to this star model, there is a *Central Node* that asks/receives information to/from both *Peripheral Nodes* and *File Repositories*. The former category is made by stakeholders that hold XML and multimedia information, whereas the latter by nodes that share only digital objects. The role of the *Central Node* is to coordinate such a multiplicity of actors, thus presenting music information to the *Web Server* as if it were contained in a single IEEE 1599 document. Similarly, the *Application Server* allows to add and remove multimedia objects, edit metadata, perform synchronizations, etc. operating on distributed objects in a transparent way. Concerning the physical location of IEEE 1599 documents, they could be hosted by the *Central Node* (centralized solution) or scattered across *Peripheral Nodes* as well, provided that the *Central Node* knows the location of the master document for each music piece. Please note that, for the sake of simplicity, this diagram deliberately ignores problems such as redundancy, network topologies with the presence of firewalls and DMZs, etc.

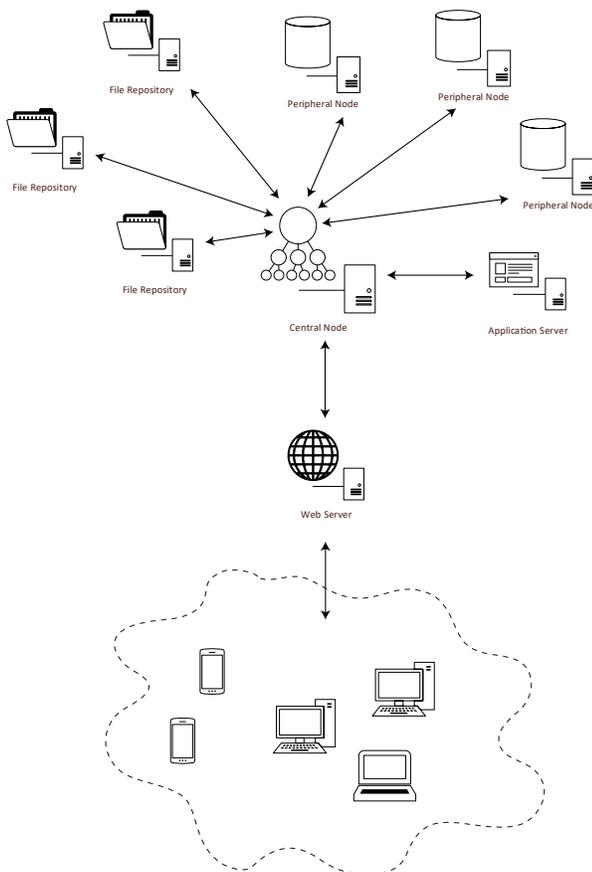


Figure 1: The proposed architecture for musical assets' stakeholders.

5 DISCUSSION AND FINAL REMARKS

The approach described so far can be mapped onto the FAIR Data Principles, a set of guidelines to make data findable, accessible, interoperable and reusable [22]. This initiative lets digital data producers and publishers promote maximum use of data. With respect to single FAIR requirements, IEEE 1599 supports rich metadata, but currently no unique and persistent identifier is associated with documents, a problem that could be easily solved by assigning DOI identifiers (*findability*); being based on XML, metadata and data are understandable to humans and machines, but there is still no trusted repository for documents (*accessibility*); metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation (*interoperability*); music data and collections can be associated with usage licenses and provide accurate information on provenance (*reusability*).

It is worth underlining that the field of multimodal analysis, that in recent years has raised more and more attention by the music research community, can benefit from the adoption of the IEEE 1599 standard is. According to [19], which extensively reviewed the Multimodal Music Information Retrieval (MIR), the concept of *modality* can be defined as a music representation digitized in a

particular place and time; multiple modalities can originate by digitizing music information in different places or times. This renewed interest in multimodal analysis was due to the occurrence of multiple factors, such as an increased computational power allowing for more complex approaches in everyday computers and the spread of machine-learning methods which are able to deal with heterogeneous kinds of data. In general, multimodal methods are useful in Computer Science where one modality is not enough for describing the whole object under study; music, indeed, is a complex process in which multiple abstraction levels cooperate. Despite the good results obtained through Multimodal MIR, some issue still remains for the field. The main bottleneck for these methods is the lack of federated datasets, which represents a limit for machine learning approaches. IEEE 1599, thanks to the possibility to provide multiple descriptions (also coming from different sources) and synchronize them, can be the answer.

Now, let us analyze the main advantages of the proposed approach from the stakeholder perspective, thus omitting the already-mentioned benefits for different categories of final users. First, stakeholders would keep the control over their assets, since metadata and digital objects are physically hosted on *Peripheral Nodes* and *File Repositories*.

Also the protection of intellectual property could be easily implemented, since each actor could decide a specific policy to grant access to its own assets; then, the *Central Node* would be in charge of filtering requests in order to meet such requirements, in accordance with the scenarios described in [7].

Is IEEE 1599 ready to support these ideas? Based on the state of the art, namely the 2008 version, the answer is only partially affirmative. For example, it is already possible to link remote files by providing a Uniform Resource Locator (URL), the General layer and multimedia-related layers allow to encode basic catalogue metadata, and a light support is offered to Digital Rights Management (DRM). Nevertheless, much remains to be done. For instance, some formats that have recently affirmed in the field of music interoperability (e.g., the revision of MusicXML in the context of the W3C Music Notation Community Group)⁶ have not been integrated yet. Similarly, IEEE 1599 offers no native support to commonly in-use international standards that describe:

- core metadata (e.g., the Dublin Core Metadata Initiative – DCMII) [21];
- traditional items catalogued by libraries (e.g., Machine Readable Cataloguing – MARC) [1];
- objects within a digital library (e.g., the Metadata Encoding and Transmission Standard – METS) [11];
- control policies for DRM (e.g., the Open Digital Rights Language – ODRL) [14].

The timing to launch this discussion is fortunate, since the IEEE 1599 standard is currently under revision, the process is expected to last 3 to 4 years, and the working group is open to any contribution from industry and academia. Thus, a collateral goal of this work is to encourage the active participation of domain experts to the revision effort.

⁶<https://www.w3.org/community/music-notation/>

REFERENCES

- [1] Henriette D Avram. 2003. Machine-readable cataloging (MARC) program. *Encyclopedia of library and information science* 3 (2003), 1712.
- [2] Denis L Baggi and Goffredo Haus. 2009. IEEE 1599: Music Encoding and Interaction. *IEEE Computer* 42, 3 (2009), 84–87.
- [3] Denis L Baggi and Goffredo Haus. 2013. *Music navigation with symbols and layers: Toward content browsing with IEEE 1599 XML encoding*. John Wiley & Sons.
- [4] Adriano Baratè, Goffredo Haus, and Luca Andrea Ludovico. 2016. A Critical Review of the IEEE 1599 Standard. *Computer Standards & Interfaces* 46 (2016), 46–51. <https://doi.org/10.1016/j.csi.2016.02.001>
- [5] Adriano Baratè, Goffredo Haus, and Luca A Ludovico. 2016. A critical review of the IEEE 1599 standard. *Computer Standards and Interfaces* 46 (2016), 46–51.
- [6] Adriano Baratè, Goffredo Haus, Luca Andrea Ludovico, and Davide Andrea Mauro. 2012. IEEE 1599 for Live Musical and Theatrical Performances. *Journal of Multimedia* 7, 2 (2012), 170–178. <https://doi.org/10.4304/jmm.7.2.170-178>
- [7] Adriano Baratè, Goffredo Haus, Luca Andrea Ludovico, and Paolo Perlasca. 2016. Managing Intellectual Property in a Music Fruition Environment - The IEEE 1599 Case Study. *IEEE MultiMedia* 23, 2 (2016), 84–94. <https://doi.org/10.1109/MMUL.2015.92>
- [8] Adriano Baratè and Luca Andrea Ludovico. 2013. IEEE 1599 Applications for Entertainment and Education. In *Music Navigation with Symbols and Layers: Toward Content Browsing with IEEE 1599 XML Encoding*, Denis Baggi and Goffredo Haus (Eds.). John Wiley and Sons, Hoboken, 115–132. <https://doi.org/10.1002/9781118494455.ch7>
- [9] Adriano Baratè and Luca Andrea Ludovico. 2016. Local and Global Semantic Networks for the Representation of Music Information. *Journal of e-Learning and Knowledge Society* 12, 4 (2016), 109–123.
- [10] Adriano Baratè, Luca Andrea Ludovico, Stavros Ntalampiras, and Giorgio Presti. 2019. *2019 International Workshop on Multilayer Music Representation and Processing (MMRP)*. IEEE Conference Publishing Services (CPS). <https://doi.org/10.1109/MMRP.2019.00001>
- [11] Linda Cantara. 2005. METS: The metadata encoding and transmission standard. *Cataloging & classification quarterly* 40, 3-4 (2005), 237–253.
- [12] Matthias Dorfer, Jan Hajič jr., Andreas Arzt, Harald Frostel, and Gerhard Widmer. 2018. Learning Audio-Sheet Music Correspondences for Cross-Modal Retrieval and Piece Identification. *Transactions of the International Society for Music Information Retrieval* 1, 1 (Sept. 2018), 22. <https://doi.org/10.5334/tismir.12>
- [13] Goffredo Haus and Maurizio Longari. 2005. A multi-layered, time-based music description approach based on XML. *Computer Music Journal* 29, 1 (2005), 70–85.
- [14] Renato Ianella. 2007. Open digital rights language (ODRL). (2007).
- [15] Katerina Kosta, Oscar F. Bandtlow, and Elaine Chew. 2018. MazurkaBL: Score-aligned Loudness, Beat, and Expressive Markings Data for 2000 Chopin Mazurka Recordings. In *Proceedings of the International Conference on Technologies for Music Notation and Representation*. Concordia University, 85–94. <https://doi.org/10.5281/zenodo.1290763>
- [16] Bochen Li, Xinzhao Liu, Karthik Dinesh, Zhiyao Duan, and Gaurav Sharma. 2019. Creating a Multitrack Classical Music Performance Dataset for Multimodal Music Analysis: Challenges, Insights, and Applications. *IEEE Transactions on Multimedia* 21, 2 (Feb. 2019), 522–535. <https://doi.org/10.1109/TMM.2018.2856090>
- [17] Adam Lindsay and Werner Kriechbaum. 1999. There's more than one way to hear it: multiple representations of music in MPEG-7. *Journal of New Music Research* 28, 4 (1999), 364–372.
- [18] Gabriel Meseguer-Brocal, Alice Cohen-Hadria, and Geoffroy Peeters. 2019. DALI: a large Dataset of synchronized Audio, Lyrics and notes, automatically created using teacher-student machine learning paradigm. *arXiv preprint arXiv:1906.10606* (2019). <https://doi.org/10.5281/zenodo.1492443> cite arxiv:1906.10606.
- [19] Federico Simonetta, Stavros Ntalampiras, and Federico Avanzini. 2019. Multimodal Music Information Processing and Retrieval: Survey and Future Challenges. In *Proceedings of 2019 International Workshop on Multilayer Music Representation and Processing*. IEEE Conference Publishing Services, 10–18. <https://doi.org/10.1109/MMRP.2019.00012>
- [20] Jacques Steyn. 2002. Framework for a music markup language. In *Proceeding of the First International IEEE Conference on Musical Application using XML (MAX2002)*. IEEE, 22–29.
- [21] Stuart L. Weibel and Traugott Koch. 2000. The Dublin core metadata initiative. *D-lib magazine* 6, 12 (2000), 1082–9873.
- [22] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3 (2016).