A critical review of the IEEE 1599 standard

Adriano Baratè, Goffredo Haus, Luca A. Ludovico*
Laboratorio di Informatica Musicale, Dipartimento di Informatica, Università degli Studi di Milano, Via Comelico, 39-20135 Milano, Italy

ARTICLE INFO
Article history:
Received 9 October 2015
Received in revised form 1 February 2016
Accepted 2 February 2016
Available online 11 February 2016

Keywords:
IEEE 1599
Music
XML
Multi-layer description

ABSTRACT
IEEE 1599 is an XML-based format standardized in 2008 by the Computer Society of the Institute of Electrical and Electronics Engineers (IEEE). The goal of this standard is to provide a comprehensive description of a music piece, supporting the encoding of heterogeneous aspects (symbolic, formal, graphical, audio, etc.) inside a unique XML document. The format presents advanced features such as a multi-layer information structuring and full synchronization among synchronizable entities. In this work we aim to conduct a critical review of IEEE 1599, not only providing a brief overview of its strength points but above all underlining those aspects that could be improved. The paper will also compare IEEE 1599 with other common formats for representing music information.

© 2016 Elsevier B.V. All rights reserved.

Contents
1. Introduction ........................................................................... 46
2. IEEE 1599 milestones .......................................................... 46
3. A short overview of the standard .......................................... 47
4. Comparison with other formats ............................................ 47
  4.1. XML-based formats .......................................................... 47
  4.2. Non-XML formats ........................................................... 48
5. Critical review ...................................................................... 48
  5.1. Score formats in the logic layer ........................................... 48
  5.2. The meaning of multiple instances ..................................... 49
  5.3. Digital rights management ............................................... 50
  5.4. Other issues ...................................................................... 50
6. Final remarks ....................................................................... 50
References .................................................................................. 50

1. Introduction
IEEE 1599 is an international standard originally conceived to provide a comprehensive description of all music information related to a single piece. The history of this format spans about 15 years, undergoing a continuous process of revision and development. Due to the standardization by the Computer Society of the Institute of Electrical and Electronics Engineers (IEEE), in 2008 the characteristics of the standard were set: any change would require an international balloting aiming at the release of a new version.

Nevertheless, in the last years the developers and researchers who adopted IEEE 1599 encountered some problems that should be addressed. The goal of this work is to point out a number of open issues recently emerged and propose a revision of some aspects of the standard, without distorting its original philosophy and approach.

The paper is organized as follows: Section 2 describes the key development stages of IEEE 1599, Section 3 presents a short overview to let a non-expert reader understand its main concepts, Section 4 provides a comparison with other music-oriented representation formats, Section 5 contains a critical review of the IEEE 1599 standard, and Section 6 provides a final discussion about the reasons for IEEE 1599 underutilization.

2. IEEE 1599 milestones

Even if the design and standardization of IEEE 1599 involved dozens of experts in the fields of music, musicology and computer
science worldwide, it was mainly conceived, designed and implemented inside two research centers, namely the Laboratorio di Informatica Musicale (LIM) of the Università degli Studi di Milano and the Scuola universitaria professionale della Svizzera italiana (SUPSI).

The first step was the establishment of the Technical Committee on Computer Generated Music, founded in 1992 in order to cover a vast interdisciplinary area spanning from Audio Signal Processing to Music composed or performed with the help of computers. Soon the activities of the committee focused on the problem of providing a comprehensive description of music information. In 2001 the IEEE Standard Association approved a recommended practice for the “Definition of a Commonly Acceptable Musical Application Using the XML Language”. It was the first formal step towards the standardization of the forthcoming IEEE 1599 standard.

In 2002 a milestone event took place at the Università degli Studi of Milan, namely the first IEEE International Conference on Musical Application using XML. In that occasion, an early version of the format — called MX — was presented to an audience of experts in the field of XML representations for music. Most ideas of the future standard were already present in MX [1].

In late 2008, as a result of the efforts of the IEEE Technical Committee on Computer Generated Music, the MX format overcame the balloting phase and it was internationally recognized by the IEEE. The name assigned to the standard was IEEE 1599.

After the standardization, the research and development activities of the working group moved to the design and implementation of off-line and on-line applications. In the former context, relevant initiatives involved institutions such as Archivio Storico Ricordi of Milan, Bach Archive of Leipzig, the Italian Ministry of Cultural Heritage and Activities [2].

In the framework of Web-oriented technologies and applications, an important initiative was EMIPIU. Born as a publicly-funded international scientific cooperation in 2011, EMIPIU is an acronym standing for Enhanced Music Interactive Platform for Internet User. The project aimed at the application of information technology to music cultural heritage, addressing in particular theaters, opera houses, museums, music initiatives, public and private institutions, conservatories, etc. The EMIPIU project was carried out by the Laboratorio di Informatica Musicale (LIM) — Università degli Studi di Milano [3] and the Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM) — Université Montpellier. The portal, which is still publicly available at http://emipiu.di.unimi.it, was implemented by Didael KTS S.r.l. [3].

In order to understand the importance that IEEE 1599 still has in the field of music and multimedia description and diffusion, latest applications based on this standard have been recently presented at the first Web Audio Conference organized by IRCAM and Mozilla in January 2015 [3].

The present work has been presented and discussed at the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR) — Music, Mind, Embodiment, which took place in Plymouth, UK on 16–19 June 2015.

3. A short overview of the standard

The key characteristics of the IEEE 1599 standard have been described in detail in many scientific publications. In addition to the official IEEE documentation, [4] discusses most of its aspects, ranging from the broad description of general concepts (e.g., multi-layer layout, common data structures and implementation of synchronization mechanisms) to specific application domains (e.g., the applicability of the standard to cultural heritage re-living and the support to Music Petri nets). Some commented examples are available in [5]. Describing the IEEE 1599 standard in detail goes beyond the purposes of this work. In this subsection we will cover only the basic concepts, for further details please refer to the mentioned bibliography.

The innovative contribution of the format is providing a comprehensive description of music and music-related materials within a unique framework. IEEE 1599 adopts an XML-based encoding and consequently it is fully compliant with the World Wide Web Consortium (W3C) specifications [19]. Comprehensiveness in music description is realized through a multi-layer environment. The XML format provides a set of rules to create strongly structured documents. IEEE 1599 implements this characteristic by arranging music and music-related contents within six layers:

- General — Music-related metadata, i.e. catalogue information about the piece;
- Logic — A score description in terms of music symbols;
- Structural — Identification of music objects and their mutual relationships;
- Notational — Graphical representations of the score;
- Performance — Computer-based descriptions and music performances through languages such as MIDI or MPEG4;
- Audio — Digital or digitized recordings, including video clips and movies.

Music events are identified in the encoding through a unique id thanks to a common data structure, known as the spine. After listing, sorting and marking music events in the spine, they can be described in different layers (e.g., the graphical aspect of a chord and its audio performance), and multiple times within a single layer (e.g., the sound of that chord in different audio performances or its graphical aspect in different scores). Consequently, the multi-layer environment provided by IEEE 1599 supports two synchronization modes:

1. Inter-layer synchronization, which takes place among contents described in different layers. Different layers are used to store heterogeneous information referring to the same music piece in a synchronized way;
2. Intra-layer synchronization, which occurs among the contents of a single layer, where homogeneous information is contained.

Coupling the aforementioned kinds of synchronization, it is possible to design and implement an advanced framework for music, whose goals could range from an advanced media experience to music-based edutainment, from cultural heritage re-living to music practice and education.

4. Comparison with other formats

4.1. XML-based formats

At the moment of writing, the two main XML-based alternatives to IEEE 1599 are represented by the Music Encoding Initiative (MEI) [6] and MusicXML [7].

In recent times W3C has launched the Music Notation Community Group, an initiative that aims to unify formats syntactically and semantically different in order to establish the guidelines for a standardized approach. The original goal was to develop and maintain format and language specifications for notated music used by
MusicXML falls short in several fundamental ways:

- MusicXML was designed as an interchange format between music applications and, by design, does not encode all document information required by a fully-featured music notation system. MusicXML has been conceived for musical documents expressed as common music notation, or a small set of alternatives. This limits the format’s capabilities for new notation, for non-Western styles, and for unconventional uses;
- MusicXML assumes a unitary rather than manifold model of a musical work, which makes basic on-line functionalities — such as version control and collaboration — cumbersome or impossible. Such a unitary model also prohibits creation of on-line scholarly editions that correlate multiple sketched versions of the same musical ideas or passages;
- MusicXML is historically tied to the MIDI standard [8], consequently the way some basic music characteristics (e.g., durations) are expressed may be difficult to understand for non-technicians.

Conversely:

- MEI provides a comprehensive and consistent model of musical notation, not influenced by legacy applications and MIDI compatibility;
- MEI is suitable for scholarly applications, as it provides explicit facilities for addressing the manifold nature of musical works. MEI was developed as an extensible format from the start, and community members are actively working on extensions for niche communities that would be difficult to incorporate in MusicXML. For example, neumatic and mensural notation are already standardized in MEI [9]. Similarly, new and alternative types of musical documents and encodings could be incorporated over time;
- MEI has also been used to encode characteristics of musical documents, musical metadata, for use in specialist databases and libraries.

IEEE 1599 presents most of the advantages of MEI: it is an extensible, free and open standard from the start, and its development was not influenced by specific applications. During the balloting phase of the standardization process, about one hundred independent international experts coming from both academia and industry evaluated the standard draft, providing helpful suggestions to improve the format.

Compared with MusicXML and MEI, IEEE 1599 intrinsically supports a stylesheet-like separation between content and appearance, which is one of the most important lessons to be learned from HTML and CSS. Besides, thanks to its multi-layer approach, not only graphical but also audio contents (and much more) can be encoded within a unique document.

4.2. Non-XML formats

Most formats for the representation of music have goals radically different from the ones of IEEE 1599, since they focus on only one layer, or a subset of layers. For instance, text-based formats usually encode only logic information (i.e., music symbols), without addressing layout and audio performance. A detailed discussion of the subject is clearly beyond the scope of this work.

However, there are other music-oriented initiatives aiming to achieve synchronization among different media streams. An example is the availability of many YouTube videos presenting score-following features, where an audio track is performed together with the advancing of a graphical score. This approach presents many relevant differences with respect to IEEE 1599 synchronization: first, the granularity is not symbol-by-symbol, but a group of measures are displayed together; besides, no interaction with content is supported other than launching and stopping the execution; finally, the combination of graphical and audio content is predefined and cannot be dynamically changed.

5. Critical review

In this section we will analyze a number of issues not explicitly addressed by IEEE 1599 specifications. In most cases, such problems could be (and actually have been) solved by adapting already available entities or forcing the meaning of the syntax. Moreover, IEEE 1599 has proved to be flexible enough to support domains and applications it was not conceived for, such as adaptive e-books [10] and live theatrical performances [11]. However, in the context of a possible future revision of the standard, more formal solutions can be found.

5.1. Score formats in the logic layer

One of the intuitions of IEEE 1599 is to adopt commonly in-use data formats for media objects such as images, audio and video. The resulting multi-layer structure is the one shown in the left part of Fig. 1, where layers are represented by blocks and media files by external shapes linked to blocks. This solution presents a number of advantages:

1. Structural and descriptive elements (XML syntax) are kept separated from media instances (external binary files);
2. Media objects, each characterized by its own features and requirements, can benefit from a wide range of commonly accepted file formats;
3. As a direct consequence of the previous item, IEEE 1599 supports huge archives of already available digital objects, with no need to encode them in XML format; and
4. The verbose way to describe information typical of XML is not applied to linked objects.

Unfortunately, a detailed analysis shows an aspect of asymmetry among layers. The Logic layer, namely the one devoted to a logic description of scores in terms of music symbols, supports only an internal XML representation following IEEE 1599 encoding rules. In other terms, it is not possible to link external files describing a score, even if other languages are clearly available. From a historical point of view, it is worth mentioning plain-text formats such as DARMS [12], GUIDO [13], Plaine and Easie Code [14]. Please refer to [15] for an in-depth analysis of a number of early codes for music representation. Narrowing the field to other XML-based formats that
Currently represent de-facto standards, please refer to Section 4. Please note that in the Logic layer we are interested in a purely logic description of music symbols (pitch, duration, articulation signs, etc.), consequently for score formats capable of encoding also page layout or graphical information these aspects would be ignored.

The original reason of this choice was the need to mark symbolic descriptions through the identifier assigned inside the spine. For (external) media objects, such as audio and images, time-based and space-based ways to locate events are available, respectively. For example, even if an audio file — say in raw format — does not support sound-event labeling, it is easy to place labels inside the IEEE 1599 Audio layer that refer to the occurrence of such an event at a given time. Similarly, as it regards graphic contents the Notational layer supports a bounding-box technique to surround occurrences of symbols. On the contrary, a score-oriented XML file format supporting identifiers is not available.

Nevertheless, in our opinion the possibility to link external contents would be highly desirable for a number of reasons:

- General approach — Supporting multiple score formats, in accordance with the approach followed in other layers, would make IEEE 1599 framework more coherent, extensible and effective;
- Completeness in symbolic description — Other initiatives, such as MEI, are undergoing continuous updates thanks to a very active community aiming to support an increasingly high number of symbols and notations;
- Interoperability and software tools — Other score formats can be generated and edited by software tools such as MakeMusic Finale and Avid Sibelius. Supporting them would make IEEE 1599 Logic layer compatible with a number of available commercial tools.

In order to solve the problem of spine identifiers, the Logic layer should implement an XML-oriented way to locate XML elements and attributes. A possible solution is adopting XPath, a query language for selecting nodes from an XML document which has been officially recognized by the World Wide Web Consortium [16].

5.2. The meaning of multiple instances

In Section 3 we mentioned the possibility to include multiple instances of homogeneous information inside a single layer. In simple terms, this means that the Audio layer can host m performances of the same piece, as well as the Notational layer can contain n score versions.

Unfortunately, an IEEE 1599 parser cannot distinguish between two situations which can both occur: multiple instances can refer to different events as well as to reifications of the same event. For the sake of clarity, some examples will be presented.

Concerning the contents of the Audio layer, the mentioned m tracks could be m distinct performances, each one characterized by its own artists (e.g., Rigoletto conducted by Richard Bonynge in 1971 and the one conducted by James Levine in 1993). Nevertheless, they could also be recordings of the same evening captured by m microphones on the stage and in the orchestra, and an IEEE 1599 document has no way to distinguish between these two cases. Similarly, the n page groupings available in the Notational layer could refer to different score editions as well as to the single parts composing a unique full score.

So far this possible source of ambiguity has been handled by not mixing the two approaches. For instance, in the multimedia project on Verdi’s Requiem, where multi-track audio recordings and multiple-angle video captures were available, the digital objects...
inside the Audio layer were all related to a single performance. Neverthe-
less, in our opinion this issue should be managed by ad hoc
aggregators or markers within the IEEE 1599 document.

5.3. Digital rights management

DRM systems are responsible for two main functions: 1) the
management of the rights, assets, parties and licenses, and 2) the
enforcement of the terms and conditions determined by the rights
holder and expressed into the licenses. Even if IEEE 1599 specifi-
cations introduce generic elements to express licenses on each linked
digital object, the issue of DRM in IEEE 1599 has never been explicitly
addressed so far. Needless to say, for a strongly multimedia-oriented
format such as IEEE 1599 this is a key aspect.

First, an accurate analysis must be conducted on current XML-
based standards, such as MPEG-21 REL and ODRL. Then, user-tailored
profiles should be defined to set the possibility to exercise the
play/display right on digital resources. This aspect is not trivial, since
users could be i) authorized, ii) not authorized, but also iii) partially
authorized, thus accessing only to a part of a digital resource (e.g. the
first 30 s of audio and the first page of a score), as well as to a low-
quality digital resources. Please note that IEEE 1599 introduces a new
type of right, namely the synchronization right occurring among the
multiple materials contained in a single document.

Other aspects, such as implementing a suitable architecture for
IEEE 1599 client–server information exchange, are application-
oriented problems which go beyond the purposes of the standard.

Early proposals and approaches to solve DRM-related problems
in IEEE 1599 have been addressed in [17], and they represent one of
our current research activities.

5.4. Other issues

When the IEEE 1599 standard was conceived, it was defined
through a Document Type Definition (DTD). A DTD is one of the
available ways to formally describe the legal building blocks of an
XML document by defining its structure as a list of elements and
attributes. Instead, currently W3C recommends XSD (XML Schema
Definition). Originally published in 2001, XSD 1.1 became a W3C Rec-
ommendation in April 2012. The most relevant features offered in
XSD and not available in native DTDs are namespace awareness and
datatypes, that is the ability to define element and attribute content
as containing values such as integers and dates rather than arbit-
rary text. As a consequence, a schema would provide more control
over a semantically correct preparation of an IEEE 1599 document.
Examples mainly concern the semantic validation of those aspects of
musical notation that cannot be solved through simple enumer-
ations or reference types. For example, the labels allowed for note
names can be defined and limited through an enumeration (a syntac-
tic element already available in DTDs), but the validation of measure
numbering would require more powerful tools. Needless to say, the
introduction of an XSD-based approach would imply a complete
revision of the standard definition and documentation.

Other minor issues concern IEEE 1599 repositories. First, a stan-
dardized compressed file format would be useful, both to reduce
redundancy in verbose XML documents (average compression rate is
higher than 90%) and to embed all materials — i.e. the XML source
and the linked digital objects — inside a unique file.

Besides, IEEE 1599 is focused on single music pieces (e.g., F.
Chopin’s Revolutionary Étude Op. 10, No. 12), and no syntax has been
reserved to group many music pieces into a single organized col-
lection (e.g., F. Chopin’s Op. 10 Études). Addressing the peculiarities
of multiple instances, librarians have poured decades of negotiation
into defining media “containers” and linking them with scores and
recordings, i.e. “instances”. For example, Functional Requirements
for Bibliographic Records (FRBR) is a conceptual entity-relationship
model developed by the International Federation of Library Associ-
ations and Institutions (IFLA) that relates user tasks of retrieval and
access in online library catalogues and bibliographic databases from
a user’s perspective [18]. As a result, this standard defines a number
of hierarchies, for instance the group of entities work — expression
— manifestation — item.

6. Final remarks

IEEE 1599 is usually considered elegant in design by the scientific
community, nevertheless it is surprisingly underutilized and conse-
quently undertested outside the precincts of its own developers.

The reasons for this lack of success do not reside only in the
problems highlighted in Section 5, but above all in the difficulty
of producing multimedia content rich enough to fully exploit the
potential of the format.

In fact, unlike other formats mentioned above, the strength of
IEEE 1599 consists in the possibility of connecting a multiplicity of
descriptions, different by type and potentially rich in number, and
with a very fine granularity. Unfortunately, the activity of identify-
ing musical events and determining their occurrence inside media
objects currently cannot be fully automated, and therefore it still
requires much manual labor.

The process of making a complete IEEE 1599 document starts
with the transcription of the piece through a suitable score editing
software. From this point of view, although the main applications on
the market (e.g., MakeMusic Finale, MuseScore and Avid Sibelius) do
not have a native export function, ad hoc IEEE 1599 plug-ins have
been released. This first step is faster if a transcription in another
format supported by such editing software (e.g., MIDI or MusicXML)
is already available. IEEE 1599 plug-ins create a basic XML file that
contains only general (metadata) and logic information (the spine
and the description of music events in terms of pitch and duration
organized by instrument and voice).

After obtaining an IEEE 1599 compliant document, digital objects
should be attached and the music events listed in the spine should
be mapped onto media content. In this perspective, some authoring
tools have been released by the developers of the format, thus allow-
ing an easier identification and mapping of the music events inside
graphic scores and audio files.

In any case, the process is rather laborious and complex. Encoding
a small–size score (about 5 pages) from scratch, including 2 graphi-
cal scores and 3 audio tracks to be mapped note by note, would take
approximatively 1 man-week. Although the encoding time could be
appreciably reduced thanks to some expedients (availability of a
transcription of the score, choice of a less fine granularity for music
events, etc.), this is a highly detrimental aspect.

Another problem unrelated to exquisitely technical aspects is the
lack of a strong community of reference (as is the case of MEI) or a
commercial subject (as is the case of MusicXML) that can sustain the
huge effort required to update the standard and develop software
tools.

References

[1] Goffredo Haus, Maurizio Longari, A multi-layered, time-based music descrip-
[2] Adriano Baraté, Luca Andrea Ludovico, IEEE 1599 applications for entertain-
ment and education, in: Denis Baggi, Goffredo Haus (Eds.), Music Navigation
With Symbols and Layers: Toward Content Browsing With IEEE 1599 XML
978-0-470-59716-3.
[3] Stefano Balda, Adriano Baraté, Goffredo Haus, Luca Andrea Ludovico, Davide
Andrea Mauro, Music-related media contents synchronized over the web: the
IEEE 1599 initiative, Proceedings of the 1st Web Audio Conference (WAC),
2015.