DigitCult | Scientific Journal on Digital Cultures

Advances and Perspectives in Web Technologies for Music Representation

Adriano Baratè, Goffredo Haus,
Luca A. Ludovico, Giorgio Presti
Laboratorio di Informatica Musicale (LIM)
Dipartimento di Informatica
Università degli Studi di Milano
Via Comelico 39, Milano, Italy

Abstract

Web technologies are quickly evolving in order to provide an increasing set of services to Internet users. Multimedia is one of the fields where the availability of high-speed networks, innovative devices and new technological approaches is noticeably arousing the interest of content producers and consumers. In this regard, it is worth mentioning new HTML5 elements to embed multimedia, recent evolutions of e-book formats towards interactivity and hypermedia, and the establishment of new W3C groups addressing specific media-related issues. Such technological advances can have a deep impact on digital cultures, as it regards both content production and dissemination on the Web. The current challenge is to combine new hardware and software technologies in order to radically improve and customize content experience. After discussing a number of approaches to music representation, this paper aims to review the most relevant Web technologies already available and to propose new Web applications addressing music in an advanced multimedia context.

Published 6 July 2016

Correspondence should be addressed to Luca A. Ludovico, Laboratorio di Informatica Musicale (LIM), Dipartimento di Informatica, Università degli Studi di Milano, Via Comelico 39 - Milano, Italy. Email: ludovico@di.unimi.it

DigitCult, Scientific Journal on Digital Cultures is an academic journal of international scope, peer-reviewed and open access, aiming to value international research and to present current debate on digital culture, technological innovation and social change. ISSN: 2531-5994. URL: http://www.digitcult.it

Copyright rests with the authors. This work is released under a Creative Commons Attribution (IT) Licence, version 3.0. For details please see http://creativecommons.org/licenses/by/3.0/it/

 doi:10.4399/97888548960931

Introduction

New digital media are rapidly changing many aspects of the political, social, economic and cultural lives of millions of people around the world. One of the key elements at the basis of this phenomenon is the availability of network technologies and the pervasive presence of the Web in our lives. The needs (Miller 1996) and researches (Barry and Lang 2001) that emerged soon after the creation of the Web led to the development of new languages and formats, specifically conceived to improve navigation experience through multimedia and to broadcast media contents.

In order to highlight the interest towards this subject, it is worth recalling that multimedia-related issues are currently being addressed not only by private companies, but also by the most important international computing societies and not-for-profit professional membership groups. Let us mention some relevant examples.

The World Wide Web Consortium (W3C) is the main international standards organization for the World Wide Web. Besides developing and maintaining open standards – such as HyperText Markup Language (HTML), Extensible Markup Language (XML), Scalable Vector Graphics (SVG), and many others – the W3C also engages in education and outreach, develops software and serves as an open forum for discussion about the Web. As explained below, there are some recent W3C initiatives explicitly addressing multimedia support, such as new HTML5 specifications, Web MIDI API and a W3C working group on music notation.

The Association for Computing Machinery (ACM) is an international society for computing. It was founded in 1947 and is currently organized into more than 150 local chapters and about 40 Special Interest Groups (SIGs), through which it conducts most of its activities. ACM groups publish a large number of specialized journals, magazines, and newsletters. Concerning multimedia, it is worth citing *Transactions on Multimedia Computing, Communications and Applications (TOMM)*. Besides, many of the SIGs sponsor regular conferences which are the dominant venue for presenting innovations in a given field (e.g., the annual conference *ACM Multimedia*).

The Institute of Electrical and Electronics Engineers (IEEE) is a professional association formed in 1963 whose objectives are the educational and technical advancement of electrical and electronic engineering, telecommunications, computer engineering and allied disciplines. One of its main activities is to design, discuss and release standards through its standardization committees. Besides, also IEEE produces scientific literature in the electrical/electronics engineering and computer science fields, publishing over 100 peer-reviewed journals (e.g., *IEEE MultiMedia*, *IEEE Transactions on Multimedia*, etc.), and sponsors several hundred annual conferences (e.g., an annual *IEEE International Symposium on Multimedia*).

The ACM and the IEEE Computer Society are both international organizations for academic and scholarly interests in computing. Unlike the IEEE, the ACM is solely dedicated to computing. Some initiatives regarding multimedia are common to the two institutions, such as the journal entitled *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*.

The fact that the largest non-profit institutions demonstrate their interest in multimedia on the Web (through standards, conferences, scientific publications, etc.) is also a clear sign of the excitement among researchers, developers and content managers, attracted by new possibilities offered by Web technologies.

Some recent achievements in this field can be considered as enabling technologies, since these innovation can be (and actually have been) applied to drive radical changes in the capabilities of technological devices, paving the way for new application domains.

The case of HTML5, a markup language for structuring and presenting content on the Web published on October 2014 by W3C, can provide a clarifying example. As stated in (Anthes 2012), HTML5 enhanced family of specifications is leading a Web revolution propelled by a proliferation of mobile devices and social networks, which translates into new power to developers and new capabilities to users. The introduction of *ad hoc* syntactic elements to support multimedia lets Web designers and programmers release Web sites equipped with audio and video features with no need of external libraries (e.g., the proprietary Adobe Flash plugin, a sort of *de facto* standard). The HTML5 innovation implies tangible effects that go far beyond the technological sphere: in the field of digital cultures, through a browser application now it is possible to remotely experience collections of audio recordings (Baratè, Haus, and Ludovico 2015a), to implement distributed

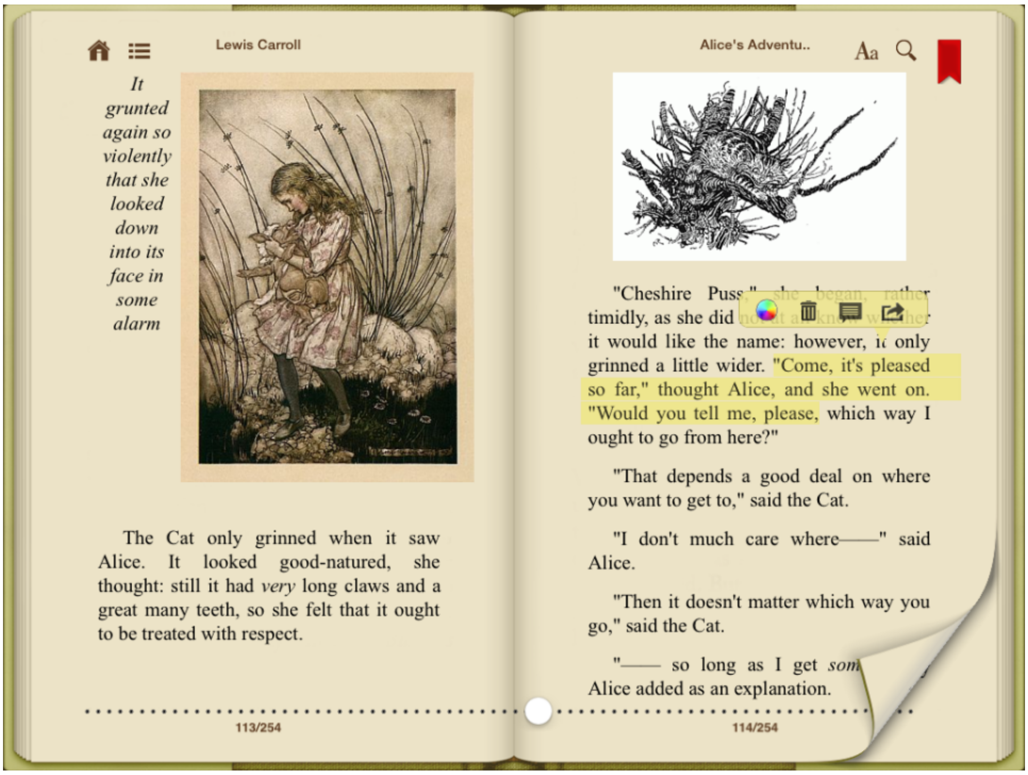

Figure 1. An example of e-book in EPUB 3 format.

music-performance systems (Allison 2011), and to release advanced tools for music learning (Ludovico and Mangione 2014).

Moreover, enabling technologies are characterized by subsequent derivative technologies. For example, the introduction of audio and video as standard elements in HTML5 has influenced other formats and products. With the release of the EPUB 3 specification (Garrish 2011), HTML5 officially becomes a part of the EPUB standard. Consequently, digital publishers are able to take full advantage of the new set of features to add “rich media” and interactivity to their e-book contents. For example, Figure 1 shows an e-book version of *Alice’s Adventures in Wonderland* by Lewis Carroll. Thanks to the EPUB 3 format, dynamic contents based on HTML5 or CSS3 are available: users can enjoy functions for adjusting fonts, colors and margins, highlight capabilities for text selection, soft and natural turning page, scroll, and enlarge effects.

The integration of HTML5 and EPUB allows a number of advanced and innovative applications, as reported in (Chesser 2011) and (Baratè, Ludovico, and Mangione 2014). A relevant example tailored to music will be discussed below.

Music can be seen as a subset of multimedia, but we will show how music contents can involve different spheres, thus providing the perfect field to test cross-domain approaches. This work aims to provide a comprehensive review of the most advanced Web technologies in the field of music, also opening a window to the future. The goal is to show both the current and the prospective impact of such technologies on digital cultures. As it regards the structure of the paper, first some methodological aspects of music description will be addressed, in order to review the most relevant approaches; then, different Web-oriented implementations will be reviewed, thus introducing the key formats currently available; finally, we will showcase some clarifying examples and identify new trends and future perspectives.

How to Structure Music Information

The description of a music work – both in real life and in the digital world – can have multiple meanings and involve different domains. A music piece usually presents an implicit or explicit definition in terms of music symbols, sometimes referred to as *logic description*. In this sense, musical notation is any system that represents scores through the use of written symbols. Many different ways are allowed for notated music, ranging from modern staff notation (Read 1979) to neumes (Parrish 1978), from Asian solmization (Kimiko and Yoshihiko 1983) to Indian *sargam* (Shirali 1977), from lute tablature (Rubsamen 1968) to Braille (World Blind Union and Krolick 1996). Once fixed such a logic description by the author, music symbols can be instanced both in the graphical domain as the result of transcription (thus producing different versions of the score), and in the audio domain as the result of interpretation (thus producing different music performances). Other kinds of information may further enrich the characterization of a given music work, for instance lyrics, on-stage photos, playbills, etc. Last but not least, let us point out the importance of text metadata that are commonly used to identify a given piece, including title (e.g., "Aida"), authors (e.g., "Ravel's Bolero"), catalog number (e.g., "Piano Sonata No. 11, K. 331"), performers (e.g., "My Way by Frank Sinatra"), etc. If we aim to provide a comprehensive description of music, all these heterogeneous aspects should be considered.

As it regards the digital domain, information entities (both data and metadata) are now in the form of bit streams, namely sequences of binary digits. In broad terms, a digital object can be originated natively in digital format – as an audio track obtained through digital sound synthesis – or it can come from a digitization process. *Digitization* is the representation of a physical object, image, sound, document or signal by generating a series of numbers that describe a discrete set of its points or samples. Properly, digitization implies conversion from the analogue to the digital domain. For instance, we call digitization the process that translates the content of a vinyl disc into a set of files. However, there are other ways to obtain a digital object, for example through a migration of data from a digital media to another. It is the case of media ripping, e.g. when the content of a CD-DA is read and saved onto a hard disk as a set of PCM[1] files, with no change in representation format. Finally, recoding is an additional way to produce a digital object, in this case through a format conversion. For example, recoding occurs when an uncompressed digital audio track (e.g. a WAV file) is converted into a lossy compressed format, typically with lower quality and smaller size (e.g. an MP3 file). The previous observations can be extended to many fields of digital culture, where digitization is a key element to preserve, query, experience and disseminate information, as evidenced by numerous digitization campaigns by important cultural institutions.

Narrowing the field to music, the problem of digitally catching and describing heterogeneity has been traditionally addressed through a number of different media formats, each one targeting a specific facet of music information. The symbolic aspects of music have been represented in binary (e.g., MakeMusic Finale, MuseScore and Avid Sibelius), text-based (e.g., ABC, GUIDO and DARMS) and XML-based (e.g., IEEE 1599, MEI and MusicXML) music-notation formats; digital images can be saved in raster formats (e.g., JPEG, PNG and TIFF) or as vector graphics (less common for music contents); sound tracks in digital audio formats, such as AIFF, MP3 and WAV; and so on. A problem emerging from the adoption of ad hoc formats is the ability to grasp and relate to each other different descriptions of the same entity: for instance, how to connect different writings – belonging to different score editions – of the same music symbols, how to connect different performances – belonging to different recordings – of the same aria, and finally how to synchronize the advancement of a cursor over a music score to a timed playback of a given audio track (*score following*).

An emerging approach, easily extensible to other fields of digital cultures, is to provide a comprehensive, integrated and synchronized description of music. In our opinion, a *multilayer structure* is adequate to treat complex and rich information by keeping contents properly organized within a unique framework. The adoption of metaphors like trees for organizing and representing hierarchical information has been already investigated in scientific literature (Lima 2014). Moreover, information should be structured in order to make internal relationships emerge across different representations, thus implementing a graph model. These concepts will be explored and defined in the following sections.

[1] Pulse-code modulation (PCM) is a method used to digitally represent sampled analog signals. It is the standard form of digital audio in computers.

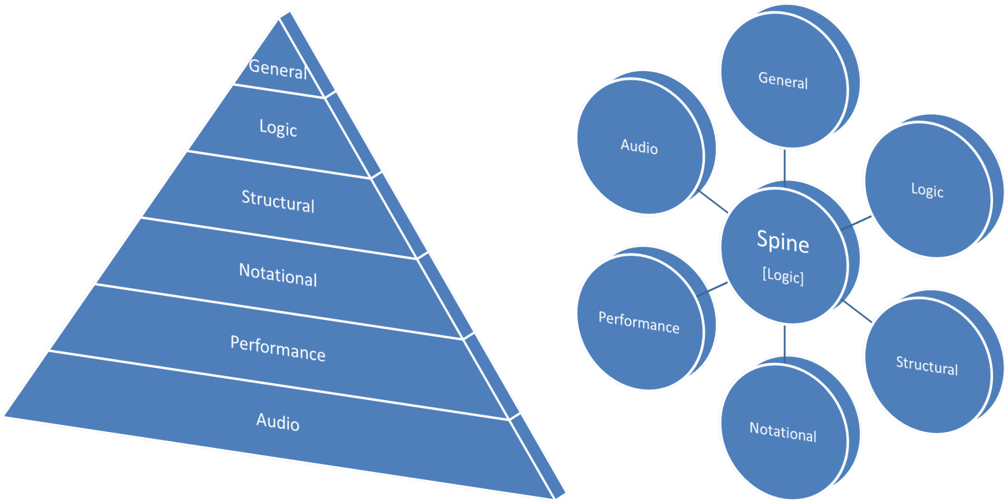

Figure 2. From a multi-layer to a star model for music representation.

The Multilayer Model

As mentioned before, our approach starts from a *multilayer model* for the description of music information. In this sense, the key concept to introduce is the idea of *layer* itself. Providing a multilayer description implies describing an entity from different perspectives, thus unveiling its heterogeneous facets. Some clarifying examples coming from different contexts are called for. A dictionary would define a layer as a covering surface which is placed onto an object, or a thickness of some material laid on or spread over a surface. In many graphics editing programs, the working area is conceived as a set of layers, where higher layers' content mask lower layers' one. In Computer Science, an abstraction layer is a way of hiding the implementation details of a particular set of functionalities. In brief, a layer enriches by adding contents and simplifying interaction, and it can be removed if lower areas have to be investigated or manipulated.

Starting from these examples, the concept of layer can be applied to music information. In fact, music information is made of heterogeneous facets whose degree of abstraction may range from a purely logical description to the physical signals.

As reported in scientific literature – e.g., see (Lindsay and Kriechbaum 1999), (Steyn 2002) and (Haus and Longari 2005) – the different aspects of music can be organized into the following layers: *General*, *Logic*, *Structural*, *Notational*, *Performance*, and *Audio*. It is worth mentioning that not all layers are necessarily present for a given music piece. Of course, the higher the number of available layers, the richer the music description. For instance, many jazz pieces have no score, since they come from extemporaneous improvisation; at most, their performance can be transcribed *a posteriori*. Similarly, many folk songs or popular tunes do not present commonly-accepted metadata or instrumentation.

The Star Model

A first refinement of the multilayer model, where each component seems to be on top of another, is represented by the *star model*.

In graph theory, a star S_n of order n, sometimes simply known as an n-star, is a tree on n nodes with one node having vertex degree $(n-1)$ and the other $(n-1)$ nodes having vertex degree 1 (Harary 1969), as shown in the right part of Figure 2.

In our context, such a structure emerges from the idea that no hierarchical relationship can be established among layers, even if they present different abstraction levels, since they provide equally-important descriptions of the same music piece. Nevertheless, a common data structure is required to keep contents organized and mutually synchronized. Such a data structure that constitute the central node is called the *spine*, since it serves as the backbone for the music description, and it is meant to univocally identify the music events to be described. In the field of music description, the concept of spine was first introduced in 1975 by David A. Gomberg, who based his proposal for music printing on a similar structure (Gomberg 1977). Even if the definition of spine (and the terminology) clearly depends on the specific format, its central role is recognized in most multilayer music-description languages.

Applying the star model to the multilayer structure mentioned above, we would obtain a star of order $n = 7$, where the $(n-1)$ leaves are the layers and the internal node is the spine. However, in literature the spine is often considered as a part of the *Logic* layer. The transformation of the unrelated multilayer model into the star model is graphically displayed in Figure 2.

In order to encompass all the music coming from different traditions, cultures, geographical areas, historical periods it is necessary to adopt a format able to be not only formal and precise in music description, but also flexible and extensible. This issue will be treated in detail in next sections.

The Graph Model

So far, the focus has been on the number of heterogeneous descriptions, which roughly corresponds to the number of supported media categories. Now it is worth underlining that each layer, say the j-th one, can contain $[1 \ldots \lambda_j]$ digital instances, and each document can be in a format appropriate to the type of media that must encode. For example, the *Audio* layer could link k WAV and MP3 audio tracks, and the *Notational* layer could refer to the digitization both of the original hand-written score in TIFF format (l files) and to m score editions, each one constituted by n_m files in JPEG and PNG formats. This kind of approach adds new leaf nodes to the star model presented above, even if from a formal point of view the former contained only layers, whereas now the model includes also digital instances (see Figure 3).

Each digital document contains a collection of music-event descriptions, which are related:

- *Internally*, when they refer to a single representation and all together constitute one of the available descriptions. It is the case of a sequence of audio events belonging to the same track (and described within the same instance of the *Audio* layer). This kind of relationship occurs among homogeneous descriptors for different music events, all parts of a single digital object;
- *Locally*, when they provide one of the various descriptions – belonging to different digital instances – of the same music event within a given layer. It is the case of the opening note of an instrumental solo notated on different scores described in the *Notational* layer. This kind of relationship occurs among homogeneous descriptors of the same music event in multiple, homogeneous digital objects;
- *Globally*, when they provide one of the various descriptions – belonging to different digital instances – retrievable from different layers for the same music event. It is the case of the opening note of an instrumental solo notated on different scores, all described in the *Notational* layer. This kind of relationship occurs among heterogeneous descriptors of the same music event in multiple, heterogeneous digital objects;

Even if the global synchronization mentioned in the last item is the only one explicitly implemented through the spine mechanism, the consequent effect is the establishment of a network of relationships among distributed descriptions of the same information entity (i.e. a given music event). On one side, explicit links can be seen as point-to-point connections between a location of a digital instance (e.g., a timing, an area, etc.) and the spine; on the other side, considering all the point-to-point links which refer to the same music event makes hidden relationships emerge. This is the core of the *graph model* for music representation.

A simplified graphical representation is shown in Figure 3, that partially illustrates the links for a single music event through continuous, dashed and dotted lines. Needless to say, an extension to all music events would create a very complex network of relationships.

Web Technologies for Music

The concept of multi-layer description – i.e. as many different types of description as possible – together with the concept of multi-instance support – i.e. as many different media objects as possible for each layer – can provide a rich and flexible way to encode music in all its aspects.

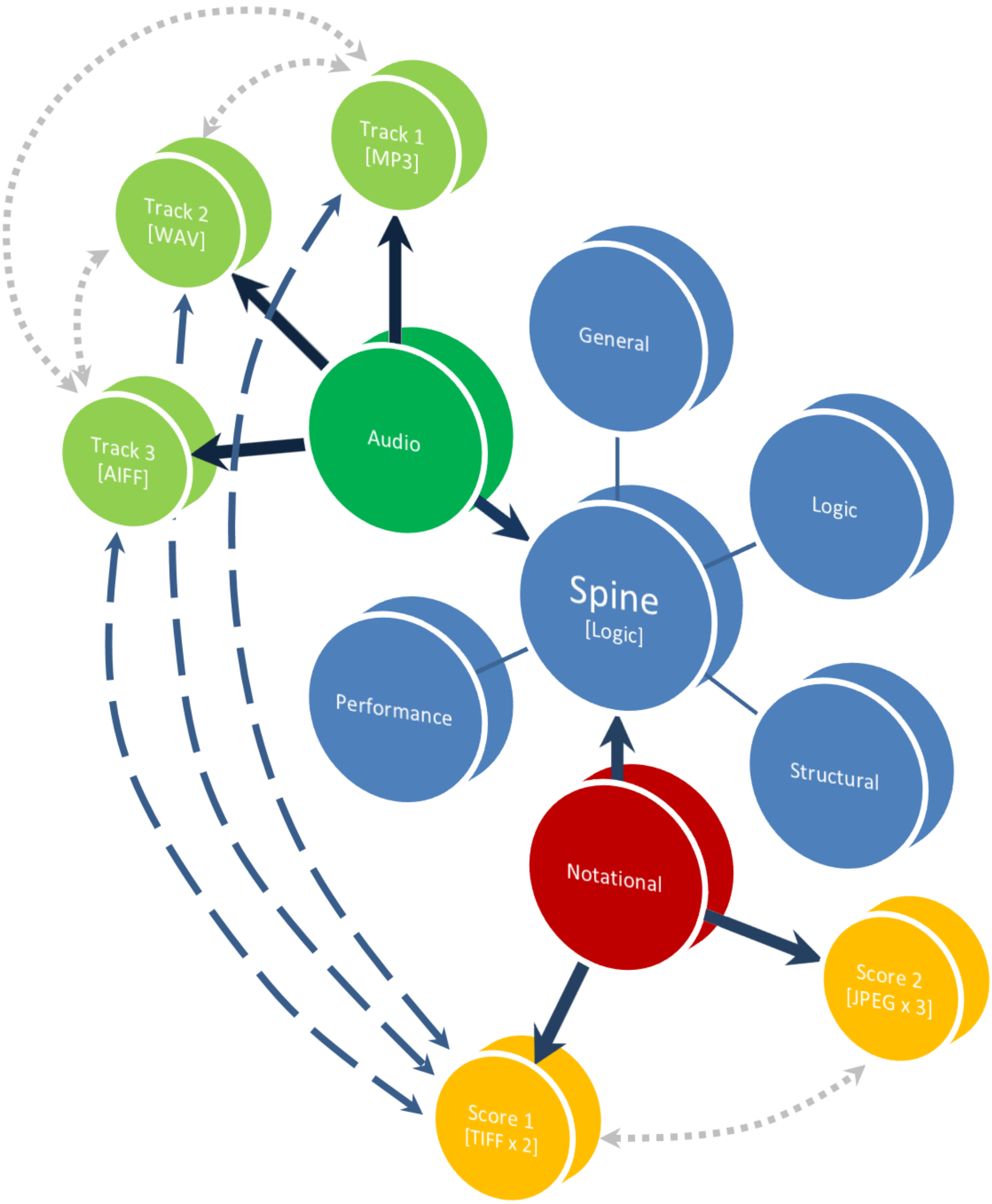

Figure 3. A graph model for music representation. Arrows represent links among different descriptions of the same music event. Continuous lines are the explicit links from instances to the spine and vice-versa, dashed lines are implicit links between instances of different layers (*inter-layer synchronization*), and dotted lines are implicit links between instances of the same layer (*intra-layer synchronization*). Implicit links are not encoded, rather they emerge when explicit links are parsed and resolved.

 doi:10.4399/97888548960931

After defining the most suitable models to structure and represent music information in a comprehensive way, now we will review the already available as well as the most promising Web technologies that can be employed to design and implement advanced music applications. A common feature is W3C compliance, which ensures accessibility from different Web browsers and devices, no matter what type of equipment technologies are accessed from.

Music Description on the Web

Music Description in XML

The concepts introduced above to support an effective and comprehensive description of music have been implemented by three relevant initiatives: MEI, MusicXML, and IEEE 1599. Since all these formats are based on XML, they can be easily integrated into Web tools and environments. XML has obvious appeal as a technology to help solve the music interchange problem. It is designed to represent complex, structured data in a standardized way. The same aspects that make XML suitable to other application areas – including straightforward usability over the Internet, ease of creating documents, and human readability – apply to musical scores as well (Haus and Longari 2002). Now these formats will be briefly described.[2]

The *Music Encoding Initiative* (MEI) is a community-driven effort to create a commonly accepted, digital, symbolic representation of music notation documents (Roland 2002). The resulting format is a set of rules for recording the intellectual and physical characteristics of music notation documents so that the information contained in them may be searched, retrieved, displayed, and exchanged in a predictable and platform-independent manner. The MEI format focuses – though not exclusively – on the encoding of documents in musicology and libraries for research and analysis purposes. Such a format is defined through a modular, extensible XML schema, accompanied by detailed documentation. Its guidelines are published under an open-source license and periodically updated. In recent releases, new elements, attributes, or content models may be included, allowing the addition of aspects that can address new types of documents. This customization approach allows the transition from a single, monolithic encoding schema to an extensible document-encoding framework. This initiative has provided not only documentation and technical specifications, but also computer-based tools widely adopted by libraries, museums, and individual scholars to encode musical scores. The reference Web site – containing documentation, guidelines, a tag library, technical details, and some encoding examples – is available at http://music-encoding.org/.

MusicXML was designed from the ground up for sharing sheet music files between applications, and for archiving sheet music files for use in the future. MusicXML has been explicitly defined by its creator as an “Internet-friendly format for sheet music”, aiming to provide a commonly-accepted interchange format (Good and Actor 2003). The goal is making MusicXML files readable and usable by a wide range of music notation applications. Originally designed and developed by Recordare LLC, it was soon integrated into leading score editors, such as Finale and Sibelius. Of the three mentioned initiatives, this is probably the most commercially successful and spread even among non-experts. Currently, MusicXML 3.1 is under development: this release will be focused on adding greater coverage for musical symbols, along with targeted bug fixes and feature enhancements. Besides, MusicXML is one of the candidate formats for W3C standardization (see next section).

IEEE 1599 is an XML-based format aiming at a comprehensive description of music information (Baggi and Haus 2013). This format has been mainly developed at the Laboratorio di Informatica Musicale (LIM) of the University of Milan, in response to IEEE Project Approval Request 1599 (P1599) - “Definition of a Commonly Acceptable Musical Application using the XML Language”. The language is a meta-representation of music information to describe, link and synchronize music-related data and metadata within a multi-layered environment. The purpose is achieving integration among symbolic, structural, notational, computer performance, and digital sound levels of representation. Furthermore, the proposed standard should integrate music representation with already defined and commonly-accepted standards. Since the core of IEEE

[2] Other simpler XML-based formats have been developed in the past, but they had neither commercial success nor scientific relevance. For a detailed review please refer to (Castan, Good, and Roland 2001).

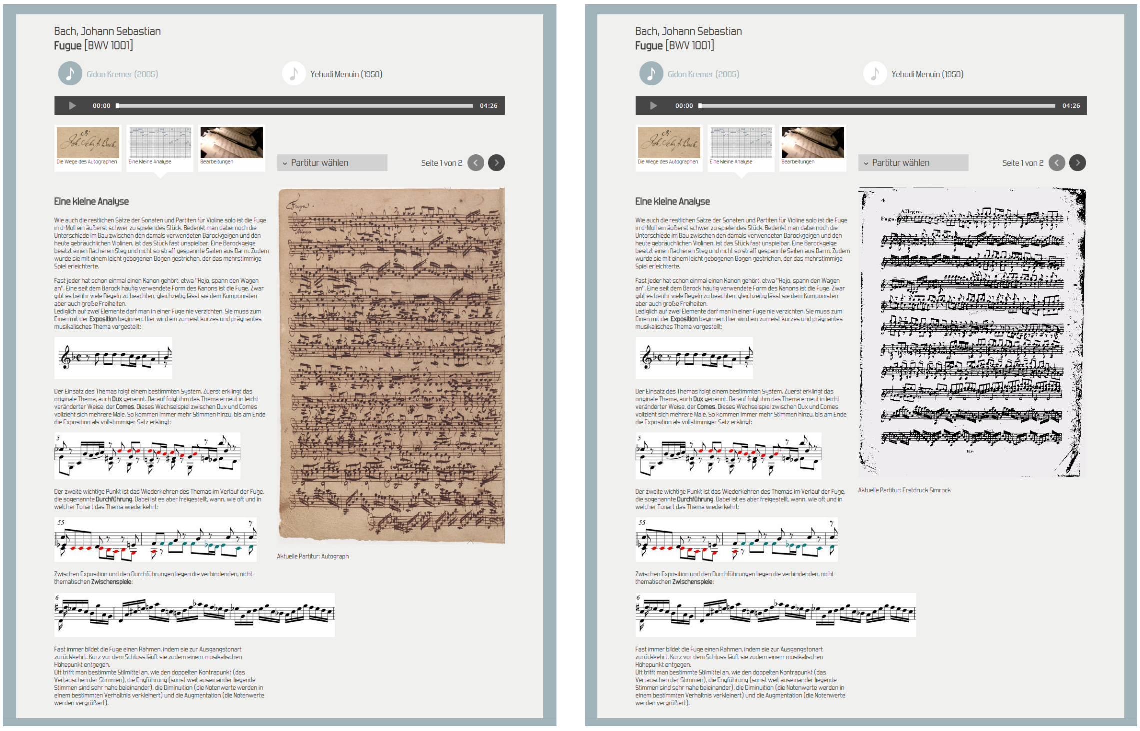

Figure 4. An IEEE 1599 interface for advanced score following, providing the user with controls to change the current audio track and score edition.

1599 is not score information, rather music symbols are considered as one of the many layers music information is made of, this is the format that best provides an extreme interpretation of content multi-layer structuring. Since September 2008, the format is an international IEEE standard. Figure 4 shows a Web interface based on IEEE 1599 and released in the context of Bach Digital project.[3]

W3C Music Notation Community Group

In recent times W3C has launched the Music Notation Community Group,[4] 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 Web, desktop, and mobile applications. The declared task of the Community Group was to maintain and update the MusicXML and SMuFL (Standard Music Font Layout) specifications, in order to evolve the specifications to handle new use cases and technologies, including greater use of music notation on the Web, while maximizing the existing investment in implementations of the existing MusicXML 3.0 and SMuFL specifications. However, the participation of independent experts (musicians, musicologists, computer researchers, software developers, etc.) has immediately fostered a more general discussion about the criteria a music notation standard must meet to have the quality, success, and longevity of *de-jure* and *de-facto* standards for other digital document representations, such as PostScript, PDF, HTML, and SVG.

At the moment of writing, the two main parties who are animating a constructive discussion are represented by MusicXML supporters on one side, and MEI enthusiasts on the other. Most experts acknowledge the influence and benefits that MusicXML has brought to the music notation community as a common language shared between applications, and they recognize the efforts for its continuous development, now conducted in an open and transparent way. However, due to the diversity of music documents and applications (e.g., composition, online critical editions, arrangement, music pedagogy, etc.), some researchers argue that MusicXML falls short in several

3 http://www.bachdigital.de/
4 https://www.w3.org/community/music-notation/

 doi:10.4399/97888548960931

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 online functionalities – such as version control and collaboration – cumbersome or impossible. Such a unitary model also prohibits creation of online scholarly editions that correlate multiple sketched versions of the same musical ideas or passages;
- MusicXML is historically tied to the MIDI standard, 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. 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.

Unfortunately, IEEE 1599 did not obtain the success and spread of other XML-based formats, and the reason must be sought among the elements that have ensured its quality and independence. In fact, MusicXML is the result of the efforts of Recordare, a private company that had business interest in developing the format and encouraging its adoption inside score editing software, whereas MEI is still relying on a fierce community of developers and users. Conversely, the IEEE initiative ended with the standardization of IEEE 1599, and right after the working group on computer generated music was closed. Consequently, the format is no more maintained and only academic research labs are keeping the initiative alive through scientific publications and the development of advanced interfaces based on this format.

Music Description in JSON

XML is not the only format for representing data in a Web environment. As it regards general purpose formats defined by open standards and specifications, it is worth mentioning the JavaScript Object Notation (JSON), a plain-text, data-interchange format based on a subset of the third edition of the ECMA-262 standard (Bray 2014).

JSON provides a mechanism for serializing data structures into strings. Its design goals were for it to be minimal, portable, textual, and a subset of JavaScript. This format can represent four primitive types (strings, numbers, booleans, and null) and two structured types (objects and arrays).

Table 1. The encoding of a measure in IEEE 1599 format and an automatic conversion into JSON, where attribute names are conventionally prefixed with @.

IEEE 1599

```xml
<?xml version="1.0" encoding="UTF-8"?>
...
<measure number="1">
  <voice voice_item_ref="guitar_voice_0">
    <chord event_ref="v0_meas1_ev0">
      <duration num="1" den="8"/>
      <notehead>
        <pitch octave="5" step="D"
                  actual_accidental="natural"/>
      </notehead>
    </chord>
    <chord event_ref="v0_meas1_ev1">
      <duration num="1" den="8"/>
      <notehead>
        <pitch octave="5" step="F"
                  actual_accidental="sharp"/>
      </notehead>
    </chord>
  </voice>
  <voice voice_item_ref="guitar_voice_1">
    <rest event_ref="v1_meas1_ev0">
      <duration num="1" den="4"/>
    </rest>
  </voice>
</measure>
```

JSON

```json
{
  "measure": {
    "@number": "1",
    "voice": [
      {
        "@voice_item_ref": "guitar_voice_0",
        "chord": [
          {
            "@event_ref": "v0_meas1_ev0",
            "duration": {
              "@num": "1",
              "@den": "8"
            },
            "notehead": {
              "pitch": {
                "@octave": "5",
                "@step": "D",
                "@actual_accidental": "natural"
              }
            }
          },
          {
            "@event_ref": "v0_meas1_ev1",
            "duration": {
              "@num": "1",
              "@den": "8"
            },
            "notehead": {
              "pitch": {
                "@octave": "5",
                "@step": "F",
                "@actual_accidental": "sharp"
              }
            }
          }
        ]
      },
      {
        "@voice_item_ref": "guitar_voice_1",
        "rest": {
          "@event_ref": "v1_meas1_ev0",
          "duration": {
            "@num": "1",
            "@den": "4"
          }
        }
      }
    ]
  }
}
```

Table 2. Disk space for score information in different formats, using one-line encoding: no white spaces, tabs and new lines (LF/CR). JSON is automatically obtained from IEEE 1599.

Piece	MusicXML	IEEE 1599	JSON
E. Satie, *Gymnopédie No. 1*	209 KB	162 KB	106 KB
J.S. Bach, *Brandenburg Concerto No. 3 – 3rd movement*	4.56 MB	4.62 MB	3.08 MB
L. van Beethoven, *Symphony No. 2 – 1st movement*	6.16 MB	3.98 MB	3.47 MB
M. Musorgskij, *Pictures at an Exhibition – Promenade*	249 KB	164 KB	102 KB
G. Verdi, *Traviata – Prelude*	1.02 MB	657 KB	389 KB

JSON provides the same functionality as XML, but it is more lightweight, since typically requires fewer characters to do so; consequently, it is sometimes defined as a "fat-free alternative to XML". An example is provided in Table 1, which compares the IEEE 1599 description of a measure and one of the many JSON translations that can be obtained through automatic processing. Table 2 shows statistical data about the disk space taken by MusicXML, IEEE 1599, and JSON encodings of the same music scores (only the logic layer).

JSON is not only efficient and effective in data representation, but also much easier to parse compared to XML. In the field of networking (e.g., Web services, client-server applications, data storing and exchange, etc.), many developers have forgone XML in favor of JSON due to its simplicity and low overhead (Ihrig 2013).

Despite the advantages that JSON seems to provide, currently there is no relevant proposal for any official music representation format. As mentioned above, experts are working on substantial improvements of already available XML-based formats, and probably JSON is seen only as a more compact, less human-readable alternative. Nevertheless, it is worth citing some individual initiatives:

- Music JSON, an early effort to convert MIDI notation into JSON format (Anentropic Blog 2009);
- JAMS, a JSON annotated music specification of reproducible MIR research (Humphrey, Salamon, Nieto, Forsyth, Bittner, and Bello 2014).

Case Studies

In this section we will present some relevant examples of already-available Web applications addressing music, and specifically involving the activities of symbolic representation, content editing/production, and integrated fruition and interaction. These case studies are meant as paradigmatic examples of what can be achieved through current technologies. For each of them, we will seek to highlight the potential impact in terms of digital cultures, also outside the field of music.

As it regards the representation of organized music symbols (i.e. scores) over the Web, the emerging of a commonly-accepted standard supported by all Web browsers is exactly the aim of the W3C working group on music notation. However, at the moment we are far from achieving that result. There are some private initiatives that aim to temporarily bridge the gap, such as the HTML5 MusicXML Viewer.[5] Let us recall that MusicXML is an XML-based open format explicitly conceived to exchange digital music. Such a Web viewer includes a library written entirely in JavaScript, able to analyze and deal with XML files in MusicXML format, and a Web-page plug-in, using the HTML5 canvas techniques to convert MusicXML documents into scores. A screenshot of the application in shown in Figure 5. Unfortunately, music notation is a very articulated field, and MusicXML presents many syntactical variants only partially implemented by score editors, so this Web prototype does not properly support rich and/or complex encodings (we unsuccessfully tested it with Grieg's *Norwegian Folk Tunes, op. 66* for piano and Beethoven's *Symphony No. 2 – 1st movement*).

Concerning the Music Encoding Initiative, there are several Web applications for visualization and analysis of scores. For example, Verovio[6] is a fast, portable and lightweight library for engraving MEI music scores into SVG. Verovio implements its own rendering engine, which can render SVG with all the musical symbols embedded in it and without dependency on external fonts. In addition to MEI, Verovio can also render MusicXML code, Plain and Easy (PAE) code and DARMS code. This software is highly modular and can be used in different contexts, including as a JavaScript toolkit, and the official Web site contains an on-line viewer. Verovio is available under the LGPLv3 license.

As it regards IEEE 1599 – whose purpose is to integrate in a unique environment many different representations of music – currently there is no Web viewer available for the visualization of scores alone. Of course it would be possible to create an *ad hoc* interface to show the content

[5] http://www.musicxml-viewer.com/
[6] http://www.verovio.org/

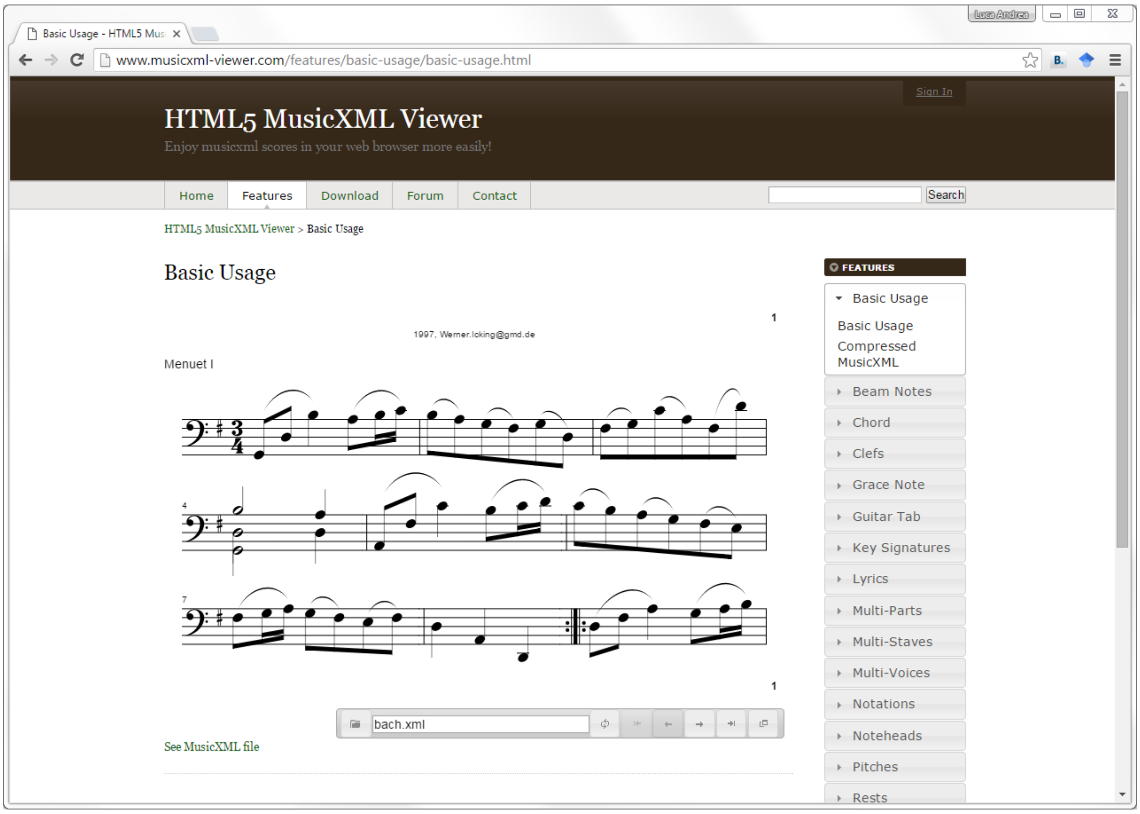

Figure 5. HTML5 MusicXML Viewer.

of the *Logic* layer, but IEEE 1599 can use (also) digitized scores to show typeset music notation, linked through the spine to the logic description music symbols. There is a number of instances of the IEEE 1599 synchronization engine, publicly available in dedicated Web sites (e.g., see the *EMIPIU* project described below) or integrated into wider projects (e.g., in the *Bach Digital* portal).[7]

Let us address a problem more general than mere score visualization. The question is: How can music technologies have an impact on digital culture? Can they influence also other research fields, other forms of expression and communication, other fruition models? The experience conducted by Pearson that brought to the release of *VideoStudio* can provide an illuminating example. *VideoStudio* is a Web interface explicitly conceived to make a traditional music textbook more interactive and adaptive. This initiative, based on the IEEE 1599 format and currently available in the Pearson Digital Library, did not simply aim to digitize the content already carried by the book (e.g., through an attached PDF version), but to promote a brand new approach to the study and teaching of music in middle school, introducing concepts such as adaptivity, interactivity, and cooperation. This Web product, shown in Figure 6, can pave the way for a new paradigm for education, and not only in the music field; for example, a possible evolution is the design and implementation of *active e-books*, as postulated in (Baratè, Ludovico, and Mangione 2014).

Finally, let us analyze the impact of Web technologies on the integrated fruition of music and the advanced user-interaction with multimedia contents. This aspect is particularly relevant, because it shows how the availability of symbolic music formats can go beyond mere representation, preservation and interchange, unveiling new music aspects and supporting new interaction modes. For instance, this is the case of *meiView*,[8] an experimental Web application designed to display 15-16th century music and provide users with a dynamic mechanism to select which variant they want to see.

Once again, an approach originally conceived for music can be pushed further, embracing multimedia in general, or even rich-information representation and fruition. For example, in the framework of the IEEE 1599 initiative, a Web portal has been published in order to make a digital

7 http://bachdigital.uni-leipzig.de/
8 http://zolaemil.github.io/meiView/

 doi:10.4399/97888548960931

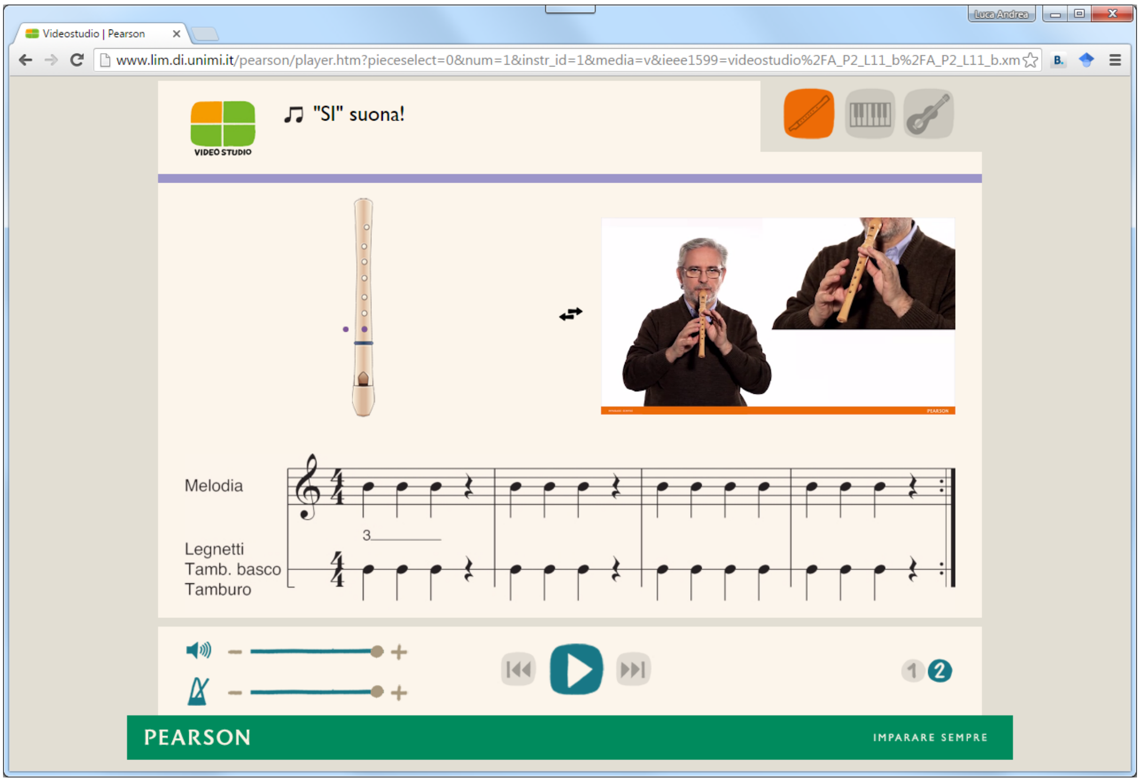

Figure 6. Pearson *VideoStudio*, a Web interface that adds digital contents and innovative features (score following, interactive score, instrument-playing guide by an expert, animations and other visual aids, etc.) to a traditional music textbook.

repository of "enriched scores" publicly available. This project, called *EMIPIU – Enriched Music Interactive Platform for Internet User)*,[9] is fully compliant with W3C standards and independent from the hardware and software characteristics of the local system in use; consequently, any device provided with a HTML5 browser and connected to the Web can access its contents. The portal includes project details, official documentation and a community area to exchange opinions, share materials and request clarifications on technical issues. However, for the purposes of this article, the most interesting section is the *Music Box*, namely a media player that implements advanced navigation and synchronization of music-related contents, as shown in Figure 7. In fact, the interface supports full synchronization among a wide variety of materials, including scores, audio and video. Moreover, for each music piece it is possible to experience different versions for any media type, e.g. different score editions or audio tracks, switching and comparing them in real time. Since the fruition of music contents occurs in a synchronized environment, this Web application implements an evolved form of score follower, too.

What Is Next?

The previous section showed a number of projects already implemented and available to the public. As mentioned above, the impact of these initiatives certainly goes beyond the scope of music, embracing more general issues that range from advanced multimedia experience to a properly-integrated representation of general-purpose information. What are the prospects in the short and medium term, in the field of music as well as in that of digital cultures?

First, we can expect a convergence of the experts towards a commonly-accepted format for the representation of music symbols over the Web. This fact would end the dispute among the supporters of different encodings, the proliferation of stand-alone software tools and plug-ins for viewing and editing, the need for format converters and translators. Consequently, the attentions and interest of musicians, musicologists, researchers and programmers could move on higher

9 http://emipiu.di.unimi.it/

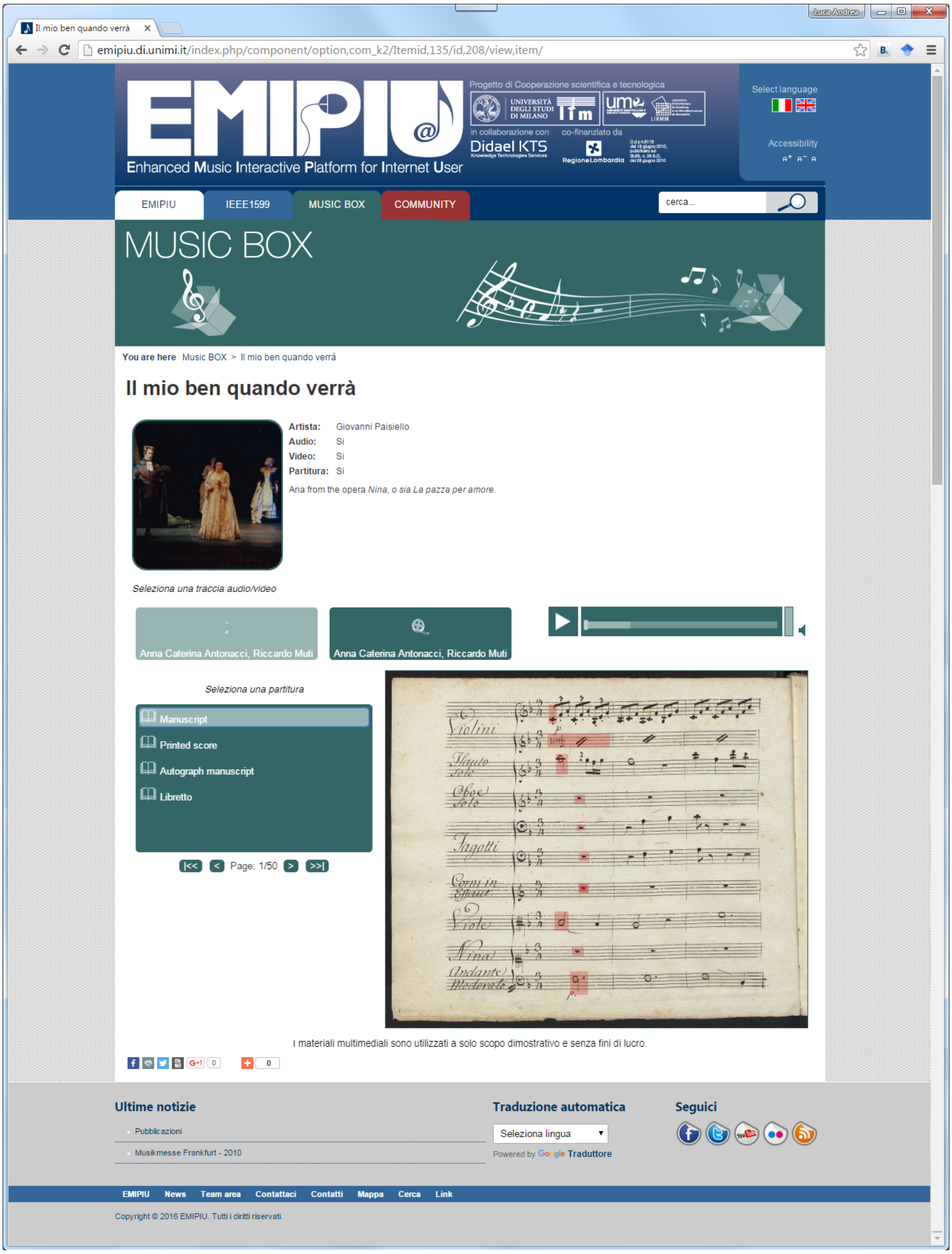

Figure 7. The *Music Box* section of the *EMIPIU* Web portal. The IEEE 1599 encoding of the current piece, namely the aria *Il mio ben quando verrà* by Giovanni Paisiello, contains 4 score versions and 2 audio/video tracks, all integrated and synchronized.

 doi:10.4399/97888548960931

goals, from the design of more stable and compatible software tools, better suited to their purposes, to the creation of a comprehensive digital-score archive in a universally recognized format, useful for the preservation, analysis and exchange of music information.

Another aspect to be considered is the increasing need for advanced tools for the production and manipulation of multimedia content. In this sense, new strategies will be based more and more on cross-media and multi-layer approaches. Early examples are already available: for instance, the *Hyperaudio* project[10] provides a set of Web tools for editing and remixing using a transcript to assemble pieces of media. Since music is a form of expression and communication that covers multiple media fields through heterogeneous channels, the expertise acquired so far can foster similar approaches in other contexts. The idea of describing an information entity from various points of view, trying to identify a plurality of descriptions and linking together its facets at various levels of granularity, can be successfully exported to other kinds of digital cultures.

Finally, wearable devices and other technologies for augmented reality are soon expected to become widespread, significantly changing the way to retrieve and access information, and above all the manner we interlink them. Probably, we can barely imagine what scenarios can be opened by the integration of a music format rich in interconnected information into the Semantic Web or the Internet of things. Possible applications include innovative services (e.g., interactive street posters and advertising), augmented experience of live music performances, non-standard and multi-modal ways to query music repositories, and so on (Baratè, Haus, and Ludovico 2015b).

In conclusion, thanks to music's inherent information structuring and multimedia features, the analysis of the advanced Web technologies for the encoding, production and enjoyment of music can be paradigmatic for understanding the future of digital cultures.

[10] http://hyperaud.io/

References

Allison, J. “Distributed performances systems using HTML5 and Rails.” In *Proceedings of the 26th Annual Conference of the Society for Electro-Acoustic Music* (2011): 1–6.

Anentropic Blog. “Music JSON: very rough outline.” https://anentropic.wordpress.com/2009/03/30/music-json-very-rough-outline/. 2009. Accessed: 2016-04-12.

Anthes, G. “HTML5 leads a web revolution.” *Communications of the ACM* 5.7 (2012): 16–17.

Baggi, D. L. and G. M. Haus. *Music Navigation with Symbols and Layers: Toward Content Browsing with IEEE 1599 XML Encoding*. Hoboken, NJ: John Wiley & Sons, 2013.

Baratè, A., G. Haus, and L. A. Ludovico. “Integration of audio resources into a digital library: The BEIC case study.” *International Journal of Digital Curation* 10.2 (2015):48–57.

Baratè, A., G. Haus, and L. A. Ludovico. “Multilayer formats and the semantic web: a music case study.” In *Proceedings of the Audio Mostly 2015 on Interaction With Sound*, edited by C. Dimoulas and G. Kalliris, 2:1–2:8. New York, NY: ACM, 2015.

Baratè, A., L. A. Ludovico, and G. R. Mangione. “An innovative paradigm for music education: Creating active e-books through the IEEE 1599 standard.” In *2014 IEEE 14th International Conference on Advanced Learning Technologies, Athens, Greece, 7-10 July 2014*, 614–616. Athens, Greece: IEEE Computer Society Conference Publishing Services (CPS), 2014.

Barry, C. and M. Lang. “A survey of multimedia and web development techniques and methodology usage.” *IEEE MultiMedia* 8.2 (2001): 52–60.

Bray, T. “The JavaScript Object Notation (JSON) data interchange format - RFC 7159.” http://tools.ietf.org/html/rfc7159.html. 2014. Accessed: 2016-04-12.

Castan, G., M. Good, and P. Roland. “Extensible markup language (XML) for music applications: An introduction.” *The Virtual Score* 12 (2001): 95–102.

Chesser, W. D. “The e-textbook revolution.” *Library technology reports* 47.8 (2011): 28–40.

Garrish, M. *What is EPUB 3?* Sebastopol, CA: O’Reilly Media Inc., 2011

Gomberg, D. A. “A computer-oriented system for music printing.” *Computers and the Humanities* 11.2 (1977): 63–80.

Good, M. and G. Actor. “Using MusicXML for file interchange.” In *Proceedings of the Third International Conference WEB Delivering of Music (WEDELMUSIC’03)*, 153. Washington DC: IEEE Computer Society, 2003.

Harary, F. *Graph theory*. Reading, MA: Addison-Wesley, 1969.

Haus, G. and M. Longari. *Proceedings of the First International IEEE Conference on Musical Application using XML (MAX2002)*. Washington DC: IEEE Computer Society, 2002.

Haus, G. and M. Longari. “A multi-layered, timebased music description approach based on XML.” *Computer Music Journal* 29.1 (2005): 70–85.

Humphrey, E. J., J. Salamon, O. Nieto, J. Forsyth, R. M. Bittner, and J. P. Bello. “Jams: a json annotated music specification for reproducible mir research.” In *Int. Society for Music Information Retrieval Conf. (ISMIR 2014)*. 2014.

Ihrig, C. J. “Javascript object notation.” In *Pro Node.js for Developers*, 263–270. New York: Apress, 2013.

Kimiko, O. and T. Yoshihiko. “Ethnomusicology in Japan since 1970.” *Yearbook for Traditional Music* 15 (1983): 155–162.

Lima, M. *The Book of Trees. Visualizing Branches of Knowledge*. New York: Princeton Architectural Press, 2014.

Lindsay, A. and W. Kriechbaum. “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.

Ludovico, L. A. and G. R. Mangione. “Teaching adaptively for music - smart opportunities emerging from the representation of score notation.” *Journal of e-Learning and Knowledge Society* 10.3 (2014): 49–67.

Miller, D. *Web multimedia development*. Berkeley, CA: New Riders Publishing, 1996.

Parrish, C. *The notation of medieval music*. Hillsdale, NY: Pendragon Press, 1978.

Read, G. *Music notation: A manual of modern practice*. Miami, FL: Taplinger Publishing Company, 1979.

Roland, P. “The music encoding initiative (MEI).” In *Proceedings of the First International Conference on Musical Applications Using XML*, 55–59. Washington DC: IEEE Computer Society, 2002.

Rubsamen, W. “The earliest French lute tablature.” *Journal of the American Musicological Society* 21.3 (1968), 286–299.

Shirali, V. *Sargam: an introduction to Indian music*. Maharashtra, India: Abhinav/Marg Publications, 1977.

Steyn, J. “Framework for a music markup language.” In *Proceeding of the First International IEEE Conference on Musical Application using XML (MAX2002)*, 22–29. Washington DC: IEEE Computer Society, 2002.

World Blind Union and B. Krolick. *New international manual of Braille music notation*. Amsterdam: SVB, 1996.