

# XML4MIR: Extensible Markup Language for Music Information Retrieval

## Abstract

### Why do we need standards?

Standards are good for scholars. They extend the scale, breadth, and accessibility of scholarly evidence and encourage innovation in learning and teaching while facilitating innovation and collaboration in scholarly discourse.

Standards are also good for business. They reduce the costs associated with acquiring and preparing data and help create new sources of revenue for that data

Bob Metcalfe has stated "the usefulness of a network grows exponentially with its number of users" (qtd. in St. Laurent). The clearest example of this statement in action is the rapid growth of the Internet. Of course, Metcalfe's Law can be applied to other technologies as well, including data encoding standards: With a standard, more users will take advantage of the technology; as the number of users increases, the price of implementing the technology falls; falling prices create competition to improve the technology; and, finally, improved technology increases the usefulness of the standard. The implications of this economic model of technology are that continuing the status quo, without a standard, is shortsighted. Both commercial (publishing) and non-commercial (scholarly) users of the technology benefit as they each function in a symbiotic relationship as both data providers and data consumers.

### Why not use existing standards?

While it might be possible to adopt an existing music representation, there are several problems with this approach.

Most existing music representations are inappropriate due to their scope. The analytical domain, which we seek to exploit in MIR, is most often the first thing to be defined as "out of scope". Many representations define their approach to music encoding too narrowly. Other representations, such as the Standardized Music Description Language, have attempted to represent music too broadly.

In addition to problems of scope, many existing solutions are hardware or software dependent or are proprietary. Therefore, their use for information exchange is severely limited.

As compromise solutions for a wide variety of materials, most meta-data systems are also inadequate for MIR. Using MARC records, it is difficult to express the complex relationships frequently found in music meta-data. TEI and EAD are designed for purposes other than music.

### Why use XML?

The foremost reason for using XML is that it is a platform-independent, open standard. Because there are no limits on the use of elements across multiple namespaces or on the structural depth that a markup language might employ, XML is very powerful. Furthermore, it is easy to implement. There is already an ever-growing set of free tools available.

The fact that the SGML/XML approach is nearing ubiquity also provides a strong reason for employing it for music representation and meta-data. Outside the music field there is a large, organized SGML/XML community already in existence, indicating a potential consumer base for encoded music materials and a pool of resources and skills that are transferable to the task of encoding music.

### Why use XML for music representation?

XML is grammar-based. XML documents can be validated using a Document Type Definition or DTD, a formal statement of the rules governing the document's grammar. There is a long history of research into the use of grammars in music description. Grammars are very powerful, they provide a generative, and utilizing them speeds encoding and reduces encoding errors.

XML is declarative. A declarative representation is preferable to a procedural one because declarative knowledge is accessible, composable, and extremely modular.

XML is structured hierarchically. When we analyze the structure of a piece of music, it is natural to name its parts and show how its elements are related. One might say XML is structurally isomorphic with music. XML's tree data structure is conceptually easy and provides efficient, non-linear data retrieval.

XML is modular. Music is often thought of as having separate visual, analytical, and performance aspects or domains. Using XML would facilitate a general encoding which allowed the facets most important to the task at hand to be represented completely and those of less importance to be minimized or left out entirely.

XML is extensible, an absolute requirement for music representation systems. A DTD can provide a mechanism for making arbitrary changes to element names and content specifications. XML is human-readable. Human-readability makes data creation and maintenance easier and functions as a protection against technological obsolescence. These considerations are very important for music because the body of material to be encoded is so vast and the investment in encoding is so large.

XML separates content and structure from presentation and behavior. Separating music representation into components, e.g. input, communication, and output, alleviates the problems of strong coupling of the syntax and semantics of the music encoding language to particular processors or processing techniques. High-level abstraction of the content allows particular representations to be generated as needed. The content can be re-used more effectively, media independence in publishing can be achieved without altering the basic data and user autonomy can be provided through user-configurable views of the data. Input can be accomplished using a wide variety of tools, such as word processors, databases, MIDI devices, graphical interfaces, or text processors. To summarize, XML provides the music community with a method for achieving interoperability of content and style, freedom from vendor control of the data, creator control of the markup syntax, and user control of the behavior of the data. Given the complexity of the task of creating a large-scale music information retrieval system, these are advantages that we cannot afford to disregard.

## Author Information

Perry Roland  
Digital Library Research Group  
University of Virginia  
pdr4h@Virginia.edu  
<http://www.people.virginia.edu/~pdr4h/>

## Suggested Readings

Barnard, David T., et al. "Hierarchical Encoding of Text: Technical Problems and SGML Solutions." *The Text Encoding Initiative: Background and Contents*. Guest Ed. Nancy Ide and Jean Véronis. Computers and the Humanities 29 (1995): 211-231. Available: <http://www.oasis-open.org/cover/barnardHier-ps.gz>. May 12, 2000.

Buxton, William, et al. "The Use of Hierarchy and Instance in a Data Structure for Computer Music." Foundations of Computer Music. Ed. Curtis Roads and John Strawn. Cambridge, MA: MIT Press, 1985. 443-466.

Dannenberg, Roger B. "Music Representation Issues, Techniques, and Systems." Computer Music Journal 17 (Fall 1993): 20-30.

Desain, Peter and Henkjan Honing. Music, Mind and Machine: Studies in Computer Music, Music Cognition and Artificial Intelligence. Amsterdam: Thesis Publishers, 1992.

Huron, David. "Design Principles in Computer-Based Representation." Computer Representations and Models in Music. Ed. Alan Marsden and Anthony Pople. New York: Academic Press, 1992. 5-39.

Roads, Curtis. "Grammars as Representations for Music." Foundations of Computer Music. Ed. Curtis Roads and John Strawn. Cambridge, MA: MIT Press, 1985. 403-442.

Roland, Perry. MusiCat DTD. Latest version available: <http://www.people.virginia.edu/~pdr4h/MusiCat/DTD/>.

Roland, Perry. "Proposed Musical Characters in Unicode (ISO/IEC 10646)." Beyond MIDI: The Handbook of Musical Codes. Ed. Eleanor Selfridge-Field. Cambridge, MA: MIT Press, 1997. 553-562. Latest version available: <http://www.lib.virginia.edu/dmmc/Music/UnicodeMusic>.

Selfridge-Field, Eleanor. "Beyond Codes: Issues in Musical Representation." Beyond MIDI: The Handbook of Musical Codes. Ed. Eleanor Selfridge-Field. Cambridge, MA: MIT Press, 1997. 565-572.