

Who, What, When, Where and Why: Introduction and Acknowledgments (First Edition)

J. Stephen Downie, PhD
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel, Champaign IL 61820 USA
+1-(217) 265-5018
jdownie@uiuc.edu

In August of 1999, I had the honour of organizing, along with Drs. Craig Nevill-Manning and David Huron, the “Exploratory Workshop on Music Information Retrieval (MIR)” at the Association for Computing Machinery’s Special Interest Group on Information Retrieval 1999 annual conference (ACM SIGIR ’99). In the workshop’s Call for Participation (http://www.lis.uiuc.edu/~jdownie/mir_papers/sigir99_wshop_proc.pdf), I set out as one of the official objectives of the meeting the following:

2. Foster a framework for future fruitful research into MIR by:
 - a. having participants explore consensus opinion on the establishment of research priorities, inter-disciplinary collaborations, evaluation standards, test collections, resource sharing, funding opportunities, communication channels, etc.
 - b. having participants begin work on the establishment of a formal organization of those interested in MIR. Such an organization will have as its mandate those items mentioned above.

The meeting was as successful as one could have hoped with the capacity crowd of participants interacting with one another in a manner that encouraged all present that interest in MIR research was more than a passing fad. Because of an over-ambitious agenda for the meeting, however, the issues outlined in Objective #2 were never formally addressed and were left for another day for resolution.

Just prior to the SIGIR workshop, I had the good fortune to meet with Dr. Don Byrd (then of University of Massachusetts, Amherst) and Matthew Dovey (University of Oxford, UK). Along with Tim Crawford (King’s College, UK), Byrd and Dovey were just beginning their NSF/JISC-funded Online Music Recognition and Searching project (OMRAS; <http://www.omras.org/>) and were keenly interested in evaluation issues. Byrd, Dovey and myself spent an entire afternoon brainstorming ideas about how to move MIR evaluation forward. The principal outcome of our efforts was the decision to organize the first International Symposium on Music Information Retrieval, with Dr. Byrd as its founding Conference Chair (ISMIR 2000; <http://ciir.cs.umass.edu/music2000>). By bringing together interested MIR researchers into one location, we hoped continue our discussions about MIR evaluation with the broader MIR research community.

In preparation for the ISMIR 2000 meeting, Tim Crawford and Don Byrd put together a briefing document that

outlined their thoughts on MIR evaluation (<http://ciir.cs.umass.edu/music2000/evaluation.html>). Of particular importance was their delineation of the Cranfield and TREC (Text REtrieval Conference; <http://trec.nist.gov>) models of evaluation that have been used for many years in the text retrieval community. In my own ISMIR 2000 briefing document (http://www.lis.uiuc.edu/~jdownie/mir_papers/downie_mir_eval.html), I outlined the key points that Dovey, Byrd and myself resolved as being central to the future of meaningful MIR evaluation:

1. A formalized set of MIR evaluation standards must become part of the MIR researcher toolkit if MIR development is to grow and flourish in its own right.
2. A set of music test databases of substantial size and varied content must be formed so that all MIR researchers can properly compare and contrast techniques under a variety of scenarios.
3. Arriving at a consensus concerning the specifics of the necessary standards and procedures will be very difficult given the wide range of disciplines from which our fellow MIR researchers are drawn.

At ISMIR 2000, an informal meeting was convened to discuss the issues surrounding MIR evaluation. The discussion was spirited and very informative. It became readily apparent that the MIR community was eager to create and participate in formal evaluations. Notwithstanding this interest, it also became evident that issues of copyright and the multi-disciplinary nature of MIR research would together pose serious difficulties that would need resolution before an evaluation framework could be set up. Access to music in today’s post-Napster climate of Intellectual Property Rights (IPR) enforcement is well known and needs no further comment. Problems posed by the multi-disciplinary approaches to MIR, however, are not so well-known and worthy of comment.

MIR researchers come from diverse backgrounds, including librarianship, musicology, computer science, audio engineering, traditional information retrieval, and business, to name but a few of their many “home disciplines”. Each research team naturally draws upon its own area of expertise and therefore MIR research and development has, for example, computer scientists working on matching algorithms, librarians on meta-data standards, audio engineers on signal processing, musicologists on representation issues, and so on. Given that each of these teams operates within the traditions and paradigms of their home disciplines, ISMIR 2000 participants were quick to point out that MIR evaluation must, in some way, incorporate, if possible, a wide variety of evaluation

standards and metrics. The specifics concerning these standards were again left to another day for resolution but it was resolved by those present that—as a starting point—the music contained within any evaluation testbeds that were to be created should consist of the following representations: a) MIDI; b) Score/Graphic Representation; and, c) Audio. These overlapping representations would thus allow meaningful comparisons, for example, between audio/signal processing approaches and symbolic approaches to music retrieval. Compiling collections of music in all three representations, of course, will be a challenge in and of itself.

Two tangential, yet significant, developments arose from the ISMIR 2000 meeting. First, Michael Fingerhut of IRCAM, Paris, set up the music-ir@ircam.fr mailing list which has helped enormously in uniting the MIR research community. Second, I made the acquaintance of Dr. Suzanne Lodato of the Andrew W. Mellon Foundation. Dr. Lodato was keenly aware of the communications and evaluation challenges facing the multi-disciplinary MIR research and development community. With Dr. Lodato's assistance, the Mellon Foundation came to support the creation of the Music Information Retrieval Annotated Bibliography Project (<http://music-ir.org>). More about the important role Dr. Lodato and the Mellon Foundation continue to play later.

Other than some lively discussions on music-ir@ircam.fr, little progress was made on the evaluation front between the conclusion of ISMIR 2000 and the convening of ISMIR 2001 at Indiana University, Bloomington (<http://ismir2001.indian.edu>). Dr. Byrd and I held another informal meeting on MIR evaluation where many of the now-familiar issues were again highlighted. Sensing a growing frustration on the part of the MIR community with the lack of progress being made, Dr. David Huron called for the drafting of a formal resolution on MIR evaluation outlining the needs and desires of the MIR community. Matthew Dovey promptly drafted “The ISMIR 2001 resolution on the need to create standardized MIR test collections, tasks, and evaluation metrics for MIR research and development” (see Appendix B; see also <http://music-ir.org/mirbib2/resolution>). On 17 October, 2001, a plenary session of ISMIR 2001 overwhelmingly endorsed the resolution which—as of 11 July, 2002—now bears the signatures of 86 members of the MIR community.

In keeping with the sense of urgency expressed by ISMIR 2001 participants, the idea was put forward that those interested in MIR evaluation should meet sometime prior to ISMIR 2002 (October 13-17, 2002; <http://ismir2002.ircam.fr>). Many at the meeting felt that convening briefly once a year at ISMIR gatherings was impeding any hope of establishing MIR evaluation frameworks. The ACM/IEEE Joint Conference on Digital Libraries 2002 (JCDL '02) conference was suggested as the venue because of its mid-summer meeting time and because many MIR researchers had had positive experiences at previous JCDL meetings. I submitted a MIR/MDL evaluation workshop proposal to William Hersh, Gary Marchionini, and Richard Futura of the JCDL '02 organizing committee which was graciously accepted (<http://www.ohsu.edu/jcdl/ws.html#W4>). The document you are now reading is the introduction to the “proceedings” of the JCDL workshop.

While preparing for the JCDL workshop, it became evident to me that a certain amount of momentum was gathering that, if possible, should be harnessed and enhanced. To this end, a supplementary panel session on MIR/MDL evaluation was proposed for ISMIR

2002 with an eye toward building upon and supplementing the outcomes of the JCDL workshop (<http://ismir2002.ircam.fr/panels.html#panel1>). Michael Fingerhut, ISMIR 2002 General Chair, and the rest of the ISMIR organizing committee are to be thanked for accepting and supporting the panel proposal. I hope that the combined results of the JCDL workshop and ISMIR 2002 panel will together make it possible to unveil a prototype MIR/MDL evaluation framework in time for ISMIR 2003.

With two MIR/MDL evaluation events to organize (and from which results would have to be published), it appeared to me that some organizational support would greatly increase the quality of both the meetings and their respective outcomes. I turned to Dr. Lodato and the Mellon Foundation for that support. Dr. Lodato and Dr. Don Waters patiently worked with me as I drafted incrementally more coherent proposals for an overarching framework designed to draw together the results of the evaluation discussions that were to take place. The text of the final proposal—which now forms the outline for this MIR/MDL evaluation frameworks project—is presented in the next chapter. I cannot thank Drs. Lodato and Waters and the Mellon Foundation enough for their generous moral and financial support of MIR/MDL research, development and evaluation.

Specific project goals and objectives can be found in the outline text, but allow me here to highlight that the project documentation, of which this document is a part, has three planned phases, or “editions”. The “first edition” is this document and comprises the project outline, Dr. Voorhees' Keynote paper on TREC evaluation scenarios for MIR/MDL evaluation, and the “White Papers” submitted by a variety of MIR researchers. The “second edition” will comprise these proceedings, the findings and conclusions of the JCDL Workshop participants, the White Paper submission for the ISMIR 2002 panel on MIR/MDL evaluation and the findings-to-date of the background research to be conducted by myself and my Graduate Assistants here at the University of Illinois (to be published October 2002). The “third or final edition” will comprise the preceding two editions, any new information gathered and a set of formal recommendations for establishing meaningful and sustainable MIR/MDL evaluation frameworks (to be published April 2003).

As can be seen by reading this introduction, many different people have played important roles in helping establish MIR/MDL evaluation frameworks. I would like to thank my indefatigable Graduate Assistant, Karen Medina, for all her hard work on both this and the MIR Bibliography projects. I would also like to thank Dr. Linda Smith for her never-ending moral and editorial support. My colleagues here in the Graduate School of Library and Information Science's, Information Science Research Lab, are also thanked most heartily. Also, my colleagues in the MIR/MDL world who have submitted White Papers, project suggestions and/or wishes of good will are all to be thanked. I know that I must have left out some that should be mentioned and I here apologize for the omission.