

Modelling Complex Multimedia Relationships in the Humanities Computing Context: Are Dublin Core and FRBR up to the Task?

J. Stephen Downie (*jdownie@uiuc.edu*)

University of Illinois at Urbana-Champaign

Allen Renear (*renear@uiuc.edu*)

University of Illinois at Urbana-Champaign

Adam Mathes (*adam@adammathes.com*)

University of Illinois at Urbana-Champaign

Karen Medina (*kmedina@alexia.lis.uiuc.edu*)

University of Illinois at Urbana-Champaign

David Dubin (*ddubin@uiuc.edu*)

University of Illinois at Urbana-Champaign

Jin Ha Lee (*jinlee1@uiuc.edu*)

University of Illinois at Urbana-Champaign

Introduction

It is now widely recognized that the creation, management, and analysis of content other than text is extremely important if the digital humanities are to deliver access to, and provide an analytical purchase on, the full range of human culture. However it is not clear to us whether the cataloguing and classification systems for digital content are up to the task. Difficulties in this area threaten to impede both the development of tools and techniques — and the production of sound theoretical results. In our paper we discuss some of these problems, focusing on *relationships* amongst the various cultural modes of expression. With the intention of convening a larger discussion of how these confusions might be remedied, we then propose directions for some clarification and improvement. However, the larger issues here are not merely terminological and resist any easy resolution.

The Problem

Within the humanities computing community it has been a commonplace that while the emphasis on representing

and analyzing textual content may be understandable, it is important to support the other kinds of content as well. We agree. The 'digital humanities' must support the full range of human cultural products: text, music, images, dance, cinema, architecture, design, and so on. At present there are many different research communities looking into the organization of, and enhanced access to, these various modes of cultural expression. There is a text retrieval community (see Baez-Yates & Ribeiro-Neto), a growing music information retrieval community (see Futrelle & Downie), an image retrieval community (see Hsin-liang & Rasmussen), and so on. Notwithstanding the real progress being made by each of these, very astonishingly little work has yet been done to comprehensively address the issue that each of these individual modes of expression interact with each other in the ordinary course of production, management and use, as well as how formats at varying level of abstraction interact within a single modality.

First, to illustrate how the modes of expression interact with each other, let us consider the *Othello* corpus. An incomplete inventory of the *Othello* corpus includes the novella by Giraldi Cinthio (1565) "upon which Shakespeare based his play" (Hunt), Shakespeare's play (1604), the operas by Rossini (1816) and Verdi (1887), Dvorak's concert overture, Op. 93 (1892), and the ballet by Lubovitch (2002). If we are going to create a digital humanities repository worthy of use by humanities scholars and their students, it is incumbent on us to build a system that can 'collocate', or gather up, all extant digital representations of *Othello*: all recordings, all scores, all movies, all choreographies, all libretti, all scripts, all set and costume designs, all critiques, and so on. To aid in this collocation, we need to clearly express the relationships between each of these things at both the specific and generic levels. On the specific level, we need to indicate that, for example, *Othello* choreographic labanotation *W* is directly based on *Othello* score *X*, which was specifically used in *Othello* movie *Y*, and also released in *Othello* soundtrack recording *Z*. On the generic level, we need to indicate that all *Othello* scores have some generic relationship to all *Othello* recordings, to all *Othello* movies, etc. in such a way that explicates that the works are all members of the *Othello* corpus.

Second, to illustrate interactions between formats within a single mode, consider only the music mode of the *Othello* corpus. For each musical realization there usually exists a symbolic score and its individual parts. These symbolic representations can, in turn, be represented in a variety of digital formats: MusicXML, TIFF, *Finale*, etc. The aural aspect of the music is represented in another variety of digital formats: WAV, MP3, Ogg Vorbis, etc. Again, complex relationships exist between the 'symbolic' and 'aural' representations at both the specific (e.g., recording *X* used score *Y*) and generic levels (e.g., a 'fakebook' score used to generate different recordings of

improvised renditions). Other potentially complex relationships exist because many of these formats can be used to generate the others. For example, a TIFF scan of the 'original' score can be fed through an Optical Music Recognition (OMR) system to create a MusicXML score file which can generate a MIDI file which then can generate any of the audio file formats. Further complicating matters, research is also underway to 'backwards' create scores from audio recordings which would capture, symbolically, the nuances of a given performance (e.g., Plumbley et al.).

Standards for Expressing Relationships Among and Within Modes

There is, of course, a body of work — standards and related research — within the cataloguing and classification communities that holds some promise for supporting the relationships described above. The Dublin Core (DC) is perhaps the most widely used within the digital humanities. IFLA's *Functional Requirements for Bibliographic Records* (FRBR) is becoming increasingly important. Work by organizations devoted to specific modalities such as the Federation Internationale des Archives du Film (FIAF)¹, and the International Association of Sound and Audiovisual Archives (IASA)², as well as work by such researchers as Martha M. Yee (moving pictures — see Yee), and Richard Smiraglia (music — see Smiraglia), etc., are also contributing insights and theory to this research domain.

Are We There Yet?

We have reviewed results from projects and analyses that suggest there is still much work to do before the functionality envisaged above is a reality. Here we describe one such project that attempts to use FRBR and the DC to support inter- and intra-modal relationships. The DC does in fact hold the most promise for representing these relationships in a way that enables computer supported exploitation for retrieval, navigation, analysis, and so on.

Ayres describes a project at MusicAustralia to use FRBR and DC to create a digital repository that explicates the complex relationships between the works, expressions, manifestations and items of a collection of music and lyrics found that:

The DC.Relation element can be used to display and support navigation between items with flat, horizontal relationships [i.e., inter-modal relationships like those between some music and its text]. However, the kinds of relationships MusicAustralia wants to expose are a combination of vertical [i.e., intra-modal relationships like those between a score and its recording] and

horizontal relationships, and rely heavily on abstract but well understood and demonstrable concepts of the Work and the Expression or version. At this stage, DC does not offer support exposure of navigational pathways that explicitly acknowledge both vertical and horizontal relationships. [Bracketed injections are ours.]

Indeed, a close look at Dublin Core format and type elements suggests that the level of precision, and subtlety required is probably not yet available there. For instance the DC type vocabulary includes such disparate things as 'sound', 'text' and 'physical object', and examples for 'sound' include 'music playback file format' and 'an audio compact disc' (DCMI Usage Board).

Next Steps: Exploring Ayres' Open Questions

Because the work of Ayres and her colleagues represents the most thorough examination of the combination of FRBR modelling and Dublin Core encoding to build a comprehensive multimodal repository, we are taking it as the starting point for our present work. The Ayres study uncovers a series of unresolved open questions associated with FRBR and the modelling of real-world multimodal information. In the Ayres case, the two modes are music (i.e., scores, recordings, etc.) and text (i.e., lyrics, poems, etc.). These two modes come together to create what we commonly consider to be 'songs'. To paraphrase Ayre's first open question:

1. Should we model as the primary work:
 - (a) the music;
 - (b) the text; or,
 - (c) the combination of text and music?

Ayres clearly illustrates that each modelling approach above clarifies a specific set of relationships between the music compositions and the texts while at the same time obscuring other relationships. The examination of this question has implications beyond the simpler music-text modelling case. For example, what are the implications when we attempt to model more complex cases (e.g., the *Othello* corpus, a Hollywood musical, etc.) with their exponentially growing relationships between text (novellas, plays, libretti, etc), music (i.e., notations, recordings, etc.), choreography (i.e., notations, video), and so on? Our paper examines this very question. We also explore the broader ramifications of Ayre's three related subsidiary open questions:

2. Should all notated and performed expressions of music [or dance, or text, etc.] be modelled as a single expression category?

3. Should expressions themselves be further modelled to include sub-categories for notated and performed expressions?
4. Should performed expressions based on particular notated expressions be modelled as expressions of expressions?

By examining these fundamental questions, we intend to encourage a long-overdue conversation within the humanities computing community. Unless our representation schemes do justice to the multidimensional complexity of cultural content in all its modes of expression, we will not realize the full potential of digital humanities repositories.

1. <<http://www.fiafnet.org/uk/>>
2. <<http://www.iasa-web.org/index.htm>>

Bibliography

Ayres, Marie-Louise. "MusicAustralia: Experiments with DC.Relation." Presented at DC-ANZ (Dublin Core in Australia and New Zealand) Conference in Canberra. February 2003. Accessed 2004-11-17. <<http://www.nla.gov.au/nla/staffpaper/2003/ayres1.html>>

Baez-Yates, R., and B. Ribeiro-Neto. *Modern information retrieval*. 1st ed. Reading, MA: Addison-Wesley, 1999.

DCMI Usage Board. *DCMI Type Vocabulary*. 2004. Accessed 2004-11-27. <<http://dublincore.org/documents/2004/06/14/dcmi-type-vocabulary/>>

Functional Requirements for Bibliographic Records (FRBR). UBCIM Publications, 19. Accessed 2004-11-27. <<http://www.ifla.org/VII/s13/frbr/frbr.htm>>

Futrelle, Joe, and J. Stephen Downie. "Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000-2002." *Journal of New Music Research* 32.2 (2003): 121-131.

Hsin-liang, Chen, and Edie M. Rasmussen. "Intellectual access to images." *Library Trends* 48.2 (1999): 291-302.

Hunt, Mary Ellen. *Review of San Francisco Ballet, "Othello". War Memorial Opera House, San Francisco, CA*. criticaldance.com, 2002. Accessed 2004-11-27. <http://www.criticaldance.com/reviews/2002/sfb-othello_020301.html>

Plumbley, M.D., S.A. Abdallah, J.P. Bello, M.E. Davies, G. Monti, and M.B. Sandler. "Automatic Music Transcription and Audio Source Separation." *Cybernetics & Systems* 33.6 (2002): 603-627.

Smiraglia, Richard. *The Nature of "a work": implications for the organization of knowledge*. Lanham, MD: Scarecrow Press, 2001.

Yee, Martha M. "What is a Work?" *The Principles and Future of AACR: Proceedings of the International Conference on the Principles and Future Development of AACR, Toronto, Ontario, Canada, October 23-25, 1997*. Ed. Jean Weihs. Ottawa: Canadian Library Association, 1998. 62-104.