An Electronic Corpus of Lute Music (ECOLM): Technological challenges and musicological possibilities

David Lewis, Tim Crawford and Michael Gale
Centre for Computational Creativity, City University, London, U.K.
d.lewis@city.ac.uk; t.t.crawford@city.ac.uk; m.gale@city.ac.uk

Background in musicology. The systematic investigation of historical repertories (such as that of the Western European lute) poses a number of problems. It is not sufficient merely to store graphical images of sources of lute music, since the tablature notation employed limits their accessibility to a handful of modern performers and scholars. It is necessary, therefore, to devise a system of encoding and presentation in order to provide access to the musical content of those sources. The process of encoding a historical document is an editorial one, demanding a considerable degree of expert knowledge in the subject domain. ECOLM is intended to allow knowledge of the lute and its repertory to be shared with users who do not have specialist academic knowledge or practical experience of the instrument or its sources.

Background in computer science. The technical challenges of ECOLM go well beyond straightforward application of database and web technology. Lute tablature shares several characteristics with the MIDI format, including a number of its disadvantages for abstract music representation. Lute music, being essentially polyphonic in structure, is not simple to encode in transcribed form since tablature notation contains no information regarding, for instance, note-spellings and voice-leading. It does, however, carry a good deal of information about performance practice embedded within the notation. Thus, it is necessary to keep the encoding as close in content to the original notation as possible.

Some pieces in the lute repertory were transmitted in many sources, although rarely in the same form; perhaps half of the surviving music is anonymous and much is misascribed. The scale of these problems when dealing with a corpus of, potentially, tens of thousands of pieces makes a clear argument for the development of 'intelligent' processing techniques such as information retrieval and computer-assisted analysis for the purpose.

Aims. The ECOLM project makes available a historically significant corpus of lute music, along with a substantial amount of musical and historical metadata, available via the World Wide Web. The corpus acts as both a reference resource and as a data set for work in computational musicology.

Main contribution. A summary of the work carried out so far within the ECOLM project in musicological and technical aspects, together with a presentation of some of the possibilities offered by new technologies for musicological investigations of the corpus and other repertories.

The considerable historical repertory of the European lute spans a period of some three centuries since it first began to be written down in the latter half of the 15th century. During the whole of this period, the chosen form of notation was usually one of the various forms of lute tablature – essentially a set of physical instructions for performance rather than an abstract representation of the desired musical sound – and thus virtually meaningless to all but a small number of modern players of the instrument. Tens of thousands of pieces have come down to us, yet only a small proportion of these have been made accessible through modern editions or recordings to non-specialist scholars, students and other musicians. Its historical importance (and, for that matter, its intrinsic musical quality) has, for this reason, been

consistently underestimated by modern musicology.

The ECOLM project, now in its second phase of funding by the UK Arts and Humanities Research Board (2001-2006), aims to make a selected but historically significant subset of the considerable body of lute music available via the World Wide Web to the academic and broader musical communities. ECOLM makes it possible for the music to be appreciated by non-experts, without the necessity of understanding tablature or the technique of the instrument and permits scholars to explore in a systematic way the interaction between the notation, the technique of the instrument itself, and music from parallel repertories. The ECOLM system provides for complete representations of the contents of historical tablature sources which can be viewed in the following modes of presentation: graphical images of the original; modern versions of the original tablature; audio playback of the music; transcriptions of the tablature into staff notation. It is supported by a sophisticated bibliographical and contextual cross-referencing apparatus which allows a wide range of investigations to be carried out by specialists and non-specialists alike.

In parallel with the basic corpus-building activities of ECOLM, a number of pilot musicological studies on the music and parallel repertories are being carried out. In each case, the aim is to explore techniques and methodologies that are appropriate to musicological investigations on an electronic corpus. The work is carried out with due attention to the criteria of traditional historical musicology, in the conviction that most, if not all, of the techniques used are also likely to find application outside the narrow domain of lute music.

The data assembled in the ECOLM project comprises two basic categories: musical data and metadata. The musical data class includes scanned images of the original manuscripts or printed sources in digital form of low to medium quality, mainly intended as a resource for validation of the encoding or of the 'editorial' decisions by the corpus-builders. The main body of musical data consists of full-text encodings of sources or substantial extracts in 'TabCode', a specially-devised ASCII representation. This can be used to provide on-screen renderings of the tablature and MIDI performances on the fly produced by server-side scripts in PHP.

The complementary metadata is maintained in an SQL-aware relational database system. This allows maximum flexibility of queries while being structured to allow the efficient investigation of special features of the corpus, such as the complex web of relationships between pieces, sources, composers, patrons, dedicatees, scribes, performers, teachers and pupils.

Content-based information retrieval techniques can be employed on the musical data and combined with metadata searches to allow analytical investigations of large amounts of music in their cultural context. While the full power of such analyses depends much on the development of appropriate algorithms and AI techniques, it also relies on the amassing of a significantly large corpus of music. Manual encoding is extremely arduous and error-prone, for obvious reasons. Recent advances in OCR techniques for music show some promise and are being adapted for large-scale capture of suitable musical source-material.

TabCode

The TabCode format

The musical content of ECOLM tablature sources is stored in TabCode format, an ASCII-based representation devised by Tim Crawford.[1] It encodes the tablature in a way that is compact and unambiguous and can be used to recreate the main features of the appearance and layout of the original notation as well as to provide data for computer-assisted processes such as transcription into conventional notation, playback or information retrieval. It is

designed to represent as completely as possible all graphical elements of the tablature in a manner that is readable both by humans and by machines. Recent extensions to the original specification allow the encoding of editorial amendments, observations and additions, ossias and alternative readings in such a way that a computer can parse the entries and give the user the option to view any of these different versions of the text.

Figures 1 and 2 show an extract from a tablature source and its encoding in TabCode. Rhythm signs are encoded by the capital initial of their American name (e.g. a quarter note, is encoded as `Q`) or, if they are beamed, by an appropriate number of `[` or `]` symbols; fret/string indicators follow directly. A full description of the TabCode format is available at http://www.ecolm.org/tabcode.

The musical encodings are stored as a field in the ECOLM database (see below). Scripts, written in PHP and executed on the fly by the web server, parse the TabCode to generate tablature on demand. Small or large extracts can be produced, ranging from individual tablature chords to complete diplomatic facsimiles. These are delivered to the client browser as graphics (PNG) embedded within an HTML page; they can for example be presented in parallel with a photographic image of the corresponding passage, where available. The encoded lute music may be converted, although currently only offline, into conventional music notation. Such automated score visualisations are useful as a guide to users unfamiliar with tablature but fall short of full transcriptions, which require a good deal of human intervention.

MIDI files can easily be generated from TabCode; although some way from ideal performances, these enable the user to get some immediate impression of the music without the need to understand the unusual notation.

Thus, a wide variety of output versions and formats may be derived from a single centralised encoding of the source. Among other benefits, this greatly facilitates verification and revision of the encodings at the data input stage. An interactive module allows immediate visual feedback of corrections in the input code by updating a visual display of a suitable tablature extract.

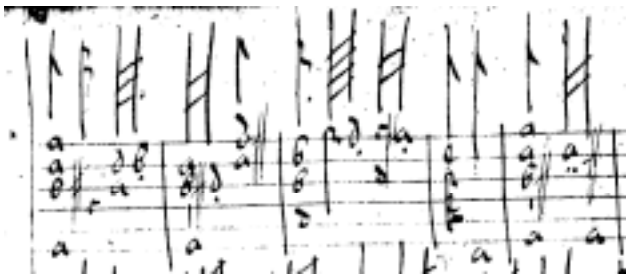

Figure 1. The opening bars from 'Lachrame mr Dowland' in US-Ws V.b.280 in tablature.

```
Qa1a2b3#a6
Ec4
[[[d2a3
]]]b2.
|
[[a2b3#(E)a6
]]d3.
Qd1#a2
|
E.b2b3d5
[[[[f1
]]]]d1.
[[c1#d3
]]a1.
|
Qe2f3e4c5
QXa
|
Qa1a2b3#(E)a6
[[a2:
]]a6
|
```

Figure 2. The opening bars from 'Lachrame mr Dowland' (see **Figure 1**) presented in TabCode.

The varied nature of tablature sources poses a number of text-critical problems for the encoder. The level of prescription varies widely between sources; some contain a good deal of extra information (ornament signs, fingering indications etc.) whereas others lack even fundamental information (e.g. rhythmic notation). Furthermore, the competence of composers, arrangers, scribes and typesetters varies enormously, meaning that the encoder must also act as an editor, providing textual alternatives

alongside a literal reading of the source. Of course, in some instances, even a literal encoding of a piece can be difficult to produce, since a number of the notational features of tablature rely on contextual factors to confirm their identity. One strength of TabCode lies in its ability to encode multiple interpretations in such instances.

TabCode, MIDI and **fret(HUMDRUM)

TabCode is a format designed for a specific type of music notation and, as such, contains information not supported in any other format (for example, it can encode information about graphical annotations or signs whose meaning is not immediately apparent). For data exchange and for computer analysis, however, we need to use more widely used standards.

As mentioned earlier, TabCode may be exported to the Standard MIDI File format, which allows audio playback within a suitably configured browser. While the deficiencies of MIDI as a general music representation are well recognised, it has clear advantages in its universality, and is well suited to the exchange of ECOLM data with many other computer-music projects.

Another significant standard comes from David Huron's *Humdrum Toolkit*, which contains tools that musicologists may find useful for analysing lute music in their own research. This suite of UNIX software tools is intended to assist in 'a wide variety of computer-based musical investigations'[2] and has a broad range of potential applications. Humdrum can import MIDI files, but lute tablature (and thus TabCode) contains performance-related information that is lost in MIDI. TabCode can be converted to **fret, a Humdrum representation designed for general tablature transcription, which allows the corpus to be used in Humdrum-based experiments.

Tablature includes a bewildering array of special signs (often unique to a source) for indicating ornaments as well as technical performance devices (fingerings, chord-spreadings, slurs, etc.). Tablature notation is furthermore the only one that directly records the idiomatic physical spacing of 'chord shapes' in the music, which strongly affects the way lutenists made arrangements ('intabulations') of music conceived for voices or for other instruments.

The performance-based, prescriptive nature of tablature notation may account for the fact that a characteristic of the manuscript repertory is that it is generally rare to encounter two versions of a piece that are identical in all details. While the essential musical 'core' or 'gist' of the work is often transmitted more or less intact, there are usually variants in interpretative features. This not only gives rise to computational problems in questions of musical similarity, but also to the need for multiple encodings which can permit the rapid comparison of parallel passages of music from different sources.

TabXML encoding of tablatures

While Tabcode is highly suitable for its purpose within the ECOLM project, it is not ideal for large-scale data exchange. For this purpose, we have started work with Dr Frans Wiering (Utrecht University) on an XML encoding scheme for tablatures, at present known as TabXML. This will permit the encoding of variants, editorial matter and other bibliographical features along the lines already established for academic texts by the *Text Encoding Initiative*[3] and in the music domain by MusicXML, MEI and other XML-based music projects.[4] The current phase of ECOLM is unlikely to use TabXML directly, but since Tabcode to TabXML conversion (and vice versa) is a fairly simple matter, it can be used for external data exchange in a variety of contexts. Early trials with the XML transformation language XSLT have demonstrated that conversions between French and Italian tablature, or between TabXML and MusicXML are certainly possible.

By the same token, data input could be carried out via XML representations. We are working with the developers of the optical symbol-recognition software *Gamera* to create an OCR system for lute tablature. The basic format for data exchange within the Gamera system is XML, and thus outputting TabXML should

be relatively straightforward. This can be converted to TabCode on the fly for ECOLM data entry.

Metadata storage and retrieval

Underlying metadata

In addition to transcriptions of musical sources, ECOLM needs to store a considerable amount of related metadata. Beyond the simple necessity of storing sufficient metadata to allow a given item to be retrieved, it was felt that a more substantial database of information would offer a number of benefits.

A principal motivation for this project was to encourage the study of a neglected and vastly under-rated body of music, so it was considered important to provide historical and geographical information, links between related sources, pieces and people, and, most importantly, bibliographical references wherever relevant. This, we hope, will help to establish ECOLM as a first point of reference for those embarking on research in the field.

A further benefit of having such metadata in machine-readable format is that it enables research-oriented retrieval: a query might, for example, easily be constructed, via the ECOLM website, to return all sources from Denmark which date from the reign of King Christian IV. Similar 'slices' of data may be taken by place, time, event, genre, form and so forth, individually or in combination.

Clearly, computational musicology and information retrieval also require 'sliced' pre-selections of data, and the benefits described above apply equally to traditional manual research and automated work. Furthermore, computational work that seeks to find associations or correlations between musical and extra-musical factors will have the necessary information directly available; for example, an algorithm for mapping changes over time in certain musical features of a genre would require the music for all examples of that genre present in the corpus and usable date information for each, all of which may be acquired with a simple database query.

In order to cater for such usage, it was considered that as much of the metadata as possible should be in computer-readable form. This, in turn, created a design dilemma: how to balance between maintaining a degree of scholarly discretion with encoding information in a purely deterministic manner for computer interpretation. It was decided, therefore, that any uncertainty or vagueness should be represented in the database wherever possible. In effect, the philosophy here is to leave as much of the interpretation of metadata to the end user as is reasonable. Such decisions have their greatest impact on how the database stores dates and personal attributions, and examples of how the database handles these will be considered below.

MySQL and XML-based standards

The selection of MySQL for the database functions of ECOLM was based on a series of factors. Although XML is clearly emerging as an industry standard for data exchange and archiving projects, SQL offers substantial advantages when one is dealing with complex webs of interrelated data. Since the current version of ECOLM exists as over 50 linked tables, it appears that the use of a relational database management system is justified. SQL also offers a powerful query language and substantial speed gains over XML.

Another factor behind the decision to use SQL was the knowledge that the generation of an XML version of the data, for example to conform with TEI, METS[5] or MODS[6] standards, should be a comparatively straightforward task.

Database structures

The ECOLM database is a complex system; for a more detailed description of its structure than can be contained in this paper, the reader is directed to the 'Database' area of http://www.ecolm.org/help/. For the

purposes of this paper, however, only a rough description is possible. Broadly speaking, the metadata can be divided into five areas: sources, pieces, people, literature and clusters.

Sources. Where possible, this will refer to a single physical entity. The source's current library location (including classmark), appearance, material and dimensions are given where available. A title is given if one appears explicitly in the source, and other names by which it is known are also stored. Linked to the source's record may be pieces, people and the nature of their involvement (with an indication of the confidence of that attribution), indications of provenance, dates, relevant bibliographical references and facsimile graphics. Pages within a source are assigned page references, with support for multiple pagination streams to indicate, for example, a source that has been both paginated and foliated.

Pieces. These are not to be confused with 'works' (see *clusters* below). A 'piece' in the ECOLM database is a single musical entity within a specified source. Multiple instances of a work, sometimes exact 'concordances', but more often closely related versions or different settings of the same basic musical material, will be counted as separate pieces. The piece is located within its source by the co-ordinates of page and system (for each page that it spans) and has a title field for its name as it appears in the source. Information about notation, instrumentation, tuning, key and form is stored with pieces as are links to dates, people (see sources above) and relevant bibliography.

A central point of reference for a user will be a list of pieces that appear within a source. Such lists are easily generated and are presented on the website with page references, facsimiles and links for more details on individual pieces.

People. People relevant to the database may have biographical information with associated places and dates stored for them, along with relevant bibliographical references. Alternative spellings of a person's name are stored in a linked table.

Literature. Stores references to any literature deemed relevant by contributors. Currently, these are all references to paper resources or links to the URLs of digital ones stored on other sites, but we plan soon to make copyright-free papers available in online form to be consulted within the ECOLM system.

Clusters. The most wide-ranging category in the database, the 'cluster' is a way of grouping any of the main data types (including clusters themselves). The use of such a group depends on the 'cluster type', the most neutral of which is 'project', which can be created by any level of 'logged-on' user as a simple way of grouping for future retrieval any material that they wish to refer to again. Such a group can be designated as private and made invisible to all other users.

More generally useful is the 'work' cluster for grouping instances of a single work in different sources. To provide more detail in such cases, hierarchical stemmatic information can be recorded using a set of clusters which distinguish levels of concordance: 'concordance', 'cognate', and so forth. Since clusters can contain other clusters, these groupings may be grouped together to give a sophisticated picture of the interrelationships between sources, between pieces, between people and between any combinations of these.

By recording alternative spellings or names for entries and by explicitly grouping pieces together in 'work' clusters, we reduce the impact of the vagaries of orthography and reliability of transmission on information retrieval. Such an approach allows, for example, a search for 'Dowland' to show works attributed to 'Dulandi' and 'Doulant' or a search for *Lachrimae* to list 'Lachramie' or even untitled works.

Handling dates. The need to strike a balance between ambiguity and

comprehensibility may be illustrated by considering how to handle historical dates in a database. In most situations encountered in historical studies, use of 'error bounds' or other such statistical delimiters is impossible. Typically, verbal or graphical indicators are used instead to show the precision with which a date is specified. Although these terms are generally clear in meaning, they have little or no intrinsic numerical significance. *Grove Music Online* provides many examples of such indicators: Tallis's dates are listed as 'b c1505; d Greenwich, 20 or 23 Nov 1585'[7], whilst other forms to be found include 'April 1377'[8], '?1565'[9] and '?1562–3'[10]. Here, dates are specified to the nearest day, month or year as is appropriate, sometimes with ranges or options, indicating inaccuracy with a 'c' and uncertainty with '?'.

To provide for the full subtlety of an English-language description of dates is impractical for this project. *TEI*, for example, permits a date to be specified as relative to another event, an immensely useful addition in many historical examples.[11] It was decided, however, that such an approach would at best greatly increase the complexity of date searching and at worst permit cases of infinite recursion (if dates are specified solely in terms of each other). The ECOLM system does, however, allow the recording of dates with null values for day, month or either of the last two digits of the year to reflect unknown values, an expression of the accuracy of each date stored as 'confident', 'probable' or 'tentative', and for that date then to be qualified as specifying a single date or a bound of a date range. For continuous processes, such as the compilation of a source, ranges may additionally be specified as referring to the start or the completion of the process.

	Thomas Tallis's death: '20 or 23 Nov 1585'		John Bull's birth: '?1562-3'	
qualifier	on	on	after	before
confidence	confident	confident	tentative	tentative
day	20	23	NULL	NULL
month	11	11	NULL	NULL
century	15	15	15	15
decade	8	8	6	6
year	5	5	2	3

Table 1. This table shows how two dates from articles in *Grove Music Online* would be encoded for the ECOLM database.

When performing advanced searches from the website, users are allowed to control the error bounds used in retrieval. This allows them to alter how the confidence value of a date should affect the search and how partial bounds should be handled. Partial bounds occur when a date such as 'after 1615' is provided with no matching 'before' date; with no cut-off point defined, a source from 'after 1615' would be retrieved even in a search for '1750' or even '2020'.

The graphical interface between the database and the client can be accessed via any web browser through the project website, http://www.ecolm.org/. Most of the information we hold is presented to the user in the form of web pages without the need to engage directly with the underlying MySQL system, although computational work can be carried out by registered users by accessing MySQL directly. Since all scripting is processed by the server, the interface is delivered in the form of pure HTML and designed to be robust to a variety of browsers of a variety of ages.

As well as serving as a reference resource, ECOLM allows distributed input and editing of data, also via the web page. Registered users can log in with a unique username and password. Depending on the level of access they are assigned, users may create and edit records and report bugs. They may also

have access to clusters and draft editions that are invisible to other users, allowing a reviewing period before new data is officially published and thus permitting group collaborations to remain private for as long as is desired.

Users may search ECOLM using the website either through conventional form fields, in which they specify their search type and terms, or by using ECOLM's 'Query builder' to build a query from (approximately) English-language phrases available as context-sensitive menu options. The resulting search may be as simple as 'Display a list of pieces the titles of which contain 'pavan'' or more complex, such as 'Display a list of sources compiled or published in London AND post-dating when William Barley commenced employment in London, UK (Publisher) AND compiled before 1615' (see figure 3). Through such a system of progressively augmented queries, sophisticated searches may be constructed in an intuitive manner.

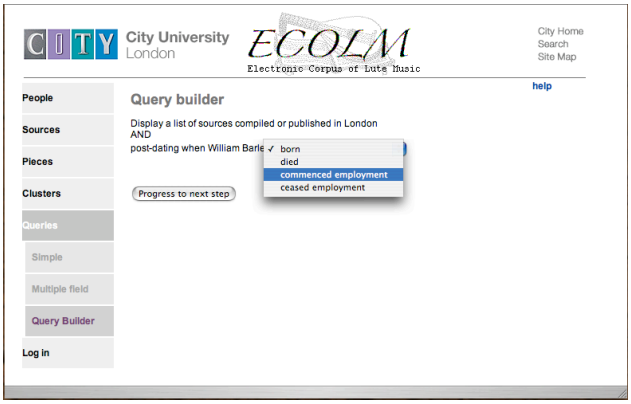

Figure 3. Screenshot showing the 'query builder' interface. This method of searching the databases allows complex searches to be constructed without knowledge of SQL or of the underlying database structure.

Projects

ECOLM may be used to support a broad range of research methodologies. We try to reflect this in the projects that we use to test the corpus, both to assess the demands on the corpus and to illustrate the variety of applications for which it may be employed.

Lachrimae. John Dowland's famous *Lachrimae* pavan is preserved in over one hundred sources in a variety of forms and scorings. Between its conception sometime in the 1580s or 1590s and the middle of the seventeenth century, *Lachrimae* was disseminated across the entire breadth of Northern Europe. We have manually collated encodings of as many versions as possible of the *Lachrimae* pavan, together with cognates and other imitations of it, in order to facilitate a large-scale comparative study. By means of a detailed textual analysis, we were able to identify a number of distinct avenues by which the work was transmitted, formulating new hypotheses regarding the genesis of a number of settings in the process.[12]

Since the *Lachrimae* are almost all built around the same framework, they present a useful data set for variation detection. Using 75 *Lachrimae* from ECOLM, mixed with around 3000 other pieces, Pickens and Crawford demonstrated that computer-based harmonic modelling could be used to search not only for instances of an individual piece, but also for related variants.[13] Further work suggests that using ECOLM metadata to divide the pieces into their formal blocks may improve retrieval still further.

Battles. 'Battle music' was widespread during the sixteenth- and seventeenth-centuries, with countless vocal and instrumental anthologies containing one or more pieces which fall into this category.

Although many pieces have their roots in Janequin's famous chanson *La guerre*, a survey of surviving 'battle music' reveals the term to be a rather loosely defined one, describing a complex network of interconnected sub-genres rather than one seminal work and its lineage.

Some of these sub-genres are quite strongly-defined (in terms of form, melodic content, etc.), but many are more flexibly conceived, taking the form of a collage of intertextual quotations and allusions, drawing from a large

collective fount of musical material. Preliminary research has shown how some of these pieces draw upon and allude to a number of other musical media in order to forge a common identity.[14] Work in progress is focussed upon the presence of a number of harmonic, melodic, motivic and textural devices (some of which stem from Janequin's chanson and its close derivatives) which also act as identifiers of the 'genre'.

Such pieces pose a serious challenge to the musicologist engaged in archival or cataloguing-based research; although the level of exact concordance between such pieces might be negligible, they nevertheless display a selection of the same characteristics and maintain a strong collective identity. These characteristics may be specified in generalised terms even though elements may be elaborated or absent.

We are developing and adapting computational tools to recognise these types of characteristic, taking into account some degree of variation. We hope to combine them to allow automated retrieval of battles from the corpus. It is hoped that such tools would be capable of being generalised for broader genre retrieval.

Further projects either currently in progress or planning include: stylometric analysis of the works of Sylvius Leopold Weiss (1687-1750) both for comparison with his contemporary, J. S. Bach, and for assessing the 'authenticity' of problematic or questionable works (the latter issue is of particular significance for the ongoing preparation of a complete edition of Weiss's music); an investigation of how audio recordings of lute music reveal aspects of historical performance practice, using ECOLM to provide 'ground-truth' editions to facilitate the computer analysis; and an examination of Lutheran chorale settings in lute sources.

Further Applications

While the ECOLM project has lute music as its primary concern, much of the technical development has equal application in other musical fields. For example, computational music-analysis techniques usually rely on a strict representation of the music's 'score' structure (note-spellings, voice-leading and so on). Much of this structure is not explicitly present in lute tablature, so special methods have to be developed to analyse music in this medium. But the same is also true of other classes of musical data, where either the structure is only partially explicit (e.g. in scores of piano music, where strict voicing is usually regarded as a lower priority than readability), or entirely lacking (as in audio recordings).

References

Wulf Arlt, 'Machaut, Guillaume de', *Grove Music Online*, ed. L. Macy (Accessed Jan 2004), http://www.grovemusic.com

Tim Crawford, 'TabCode for Lute Repertories', *Computing in musicology*, 7 (1991), 57-9

Paul Doe and David Allinson, 'Tallis, Thomas', *Grove Music Online*, ed. L. Macy (Accessed Jan 2004), http://www.grovemusic.com

Michael Gale and Tim Crawford, 'John Dowland's 'Lachrimae' at Home and Abroad', (forthcoming)

Michael Gale, 'Remnants of some late Sixteenth-Century Trumpet Ensemble Music', *Historic Brass Society Journal* 14, (2002), 115-131

David Huron, *The Humdrum Toolkit*, (CCARH: Menlo Park, CA, 1994)

Susi Jeans and O. W. Neighbour, 'Bull, John', *Grove Music Online*, ed. L. Macy (Accessed Jan 2004), http://www.grovemusic.com

Miriam Miller and Jeremy L. Smith, 'Barley, William', *Grove Music Online*, ed. L. Macy (Accessed Jan 2004), http://www.grovemusic.com

Jeremy Pickens and Tim Crawford, 'Harmonic Models for Polyphonic Music Retrieval', *Eleventh International Conference on Information and Knowledge Management*, (McClean, VA, 2002)

C. M. Sperberg-McQueen and Lou Burnard (eds.), 'TEI P4: Guidelines for text encoding and interchange', (TEI

Consortium, 2001), XML version at http://www.tei-c.org/P4X/

---

[1] Crawford (1991)

[2] Huron (1994) p.3

[3] http://www.tei-c.org/

[4] MusicXML: http://www.recordare.com/xml.html; MEI: http://dl.lib.virginia.edu/bin/dtd/mei/

[5] http://www.loc.gov/mets

[6] http://www.loc.gov/mods

[7] Doe and Allinson (accessed Jan 2004)

[8] Arlt (accessed Jan 2004)

[9] Miller and Smith (accessed Jan 2004)

[10] Jeans and Neighbour (accessed Jan 2004)

[11] Sperberg-McQueen and Burnard, http://www.tei-c.org/P4X/ND.html#nddater (section 20.4.2)

[12] Gale and Crawford (forthcoming)

[13] Pickens and Crawford (2002)

[14] Gale (2002)