

Gesellschaft für Musikforschung – Weimar, 20 September 2004

Symposion B-10: Musikalisches Erbe im digitalen Zeitalter. Chancen und Probleme neuer Techniken. Leiter: Joachim Veit

Eleanor Selfridge-Field:

The *MuseData* Electronic Corpora

When the Center for Computer Assisted Research in the Humanities was established in 1984, one of its principal goals was to form electronic corpora of musical repertories from the seventeenth, eighteenth, and nineteenth centuries.

Only people with a big appetite for unsolved problems could be attracted to such a goal. There was no MIDI or internet; email was prohibitively expensive; and personal computers, although known, were not yet in widespread use. Even in the more modest field of text-processing, ASCII was not fully accepted and systems which could print the standard diacriticals of European languages were inordinately costly and unwieldy.

We took our inspiration from the example of classicists, who over the preceding seven or eight years had encoded most surviving Greek literature from antiquity. They did it by developing their own operating system (IBYCUS), software, encoding schemes, and print drivers. By 1986 there was a computer (developed by David Woodley Packard) dedicated to the production of texts in Greek, Hebrew, and Coptic scripts as well as the Roman alphabet (including modern European diacriticals). While we did not use this computer, we did use a variant of its operating system developed by Walter Hewlett, the Center's founder. It was particularly well suited to "string" processing (i.e., the kind of data-searching people frequently do in literature).

Music required the development of many additional capabilities. The most primary was the need to print music, because proofreading remains one of the most necessary and laborious elements in the processing of corpus formation. In order to print music, it is necessary first to encode it in some manner. There are many ways to do this, none of them perfect or complete, for there is no defined number of musical symbols and many that exist derive their interpretation at least in part from their visual contextual. This led to a study of musical representation systems for the 20 years prior to the establishment of the Center and resulted, with the encouragement of the IMS Study Group on Musical Data, in the book *Beyond MIDI* (1997).

From the beginning, we have been sensitive to the problems of source *philology* and the opportunities that computers afford for enabling the presentation of multiple versions of a musical work. Projects which have put down their roots in recent years, such as those presented in this symposium, have been able to carry these aims further than we have. Our data is encoded in ASCII and each line of data provides space for annotations. However, building software to overlay annotations on a score seen on a screen remains a project for the future.

In our view of the universe, musical data which represents works of the distant past should be regarded as a springboard for virtual editions. If we look back on the mindset of a century ago, when the preparation of many *Gesamtausgaben* was vigorously pursued and the quest for an *Urtext* was strong, the trail between “popular” editions and scholarly editions seems to have been of little interest. Over a century of editing music for scholarly use, we have learned that many matters bear on the interpretation and adjudication of specific editorial issues. What we at our Center believe, however, is that whatever the dominant idea of a “best” edition may be in our own time, it is likely to be continually modified in the future. Therefore, we believe, there is a clear *philosophical* advantage in encoding music, which can be re-edited repeatedly. This attitude may help to explain why we have been slow to provide user tools for the use of our data.

As our data holdings grow, it becomes more efficient for others to invest the substantial time and expertise required to develop specific applications for them. Currently the corpora contain more than 1,000 works, chiefly from the eighteenth century—

- **Bach:** most of the instrumental music, the masses, motets, and oratorios, and many of the cantatas;
- **Corelli:** the sonatas and concerti grossi of the six published opuses;
- **Beethoven:** all of the symphonies, the violin concerto, and several quartets;
- **Haendel:** the printed opuses of instrumental music (3, 4, 6, 7); ten operas and oratorios including *Messiah* (nine versions)
- **Haydn:** the later symphonies, most of the quartets;
- **Mozart:** the later symphonies, some chamber music;
- **Telemann:** c. 150 cantatas; several dozen instrumental pieces; two major works
- **Vivaldi:** five printed opuses (1-3, 8, 10) of string and wind music; the oratorio *Juditha triumphans*.

Each of these corpora has been put to different uses by different constituencies. I cite just three of them here:

- **Haendel:** The operas and oratorios are all newly edited. All have been performed in public from *Partituren*, short scores, and parts produced from the *MuseData* databases. Most have been recorded. *Messiah* was recorded with sufficient variant movements to enable users to program nine historical performances of the work.
- **Telemann:** In the late 1980s we had a collaboration with the Telemann Zentrum in Magdeburg. Although we discontinued our own work on this project after reunification, the Zentrum has been well able to carry on.
- **Vivaldi:** Our Vivaldi scores have formed the basis of three volumes of concerto editions from Dover Publications, Inc. (New York) and more are on the way. Dover is an unusual publisher because it does not provide performing materials. Performers materials for the Dover editions are available on our website.

What are the biggest challenges facing those who are preparing electronic corpora? In a word, they are the challenges, which we as musicologists are not training to deal with. They are distribution issues, copyright issues, and user interfaces. There are some dependencies between them, but they will be ignored here.

Distribution

Distribution issues get more complicated every day. They are complicated by the multiplicity of ways in which data can be conveyed. We permit registered users to download up to 100 files a day. One hundred files may amount to only two symphonies or half an act of an opera. (Users undertaking serious projects with large amounts of data are welcome to request the data on a CD-ROM). We have found that without some kind of control, however, commercial robots and hackers can cause many problems on our website.

As users of other people's data and products, we often find well-intentioned restrictions quite disabling. Whatever the path to finding a happy medium between control and ease of access may be, we doubt it lies within the control of musicologists.

Copyright

Copyright is, of course, a much more fundamental and complicated issue. CCARH has been studying music-copyright issues almost since its formation. Regrettably, we only encode out-of-copyright editions, such as the *Bach Gesellschaft*, because no permission has been forthcoming to do otherwise. We wonder, however, whether work on current *Gesamtausgaben* will be ignored by future generations. Although much of it is prepared on a computer, the additional value of the searchable and modifiable computer materials which could form the basis of new kinds of research tools and editions yet to come has not been recognized. In fact, some publishers systematically destroy such data after production as a safeguard against misappropriation.

In European art music the differences in copyright provisions between countries, not to mention general differences between Europe and the US, perplex the most knowledgeable of legal scholars. The table below calls attention to a few outstanding differences between US and European copyright provisions related to music.

EUROPE	US
UNITED KINGDOM: Recognizes graphical-image copyright to be separate from content copyright	No recognition of separate rights for graphical image (e.g., a page of music with a particular layout, size, fonts, etc.)
UNITED KINGDOM: Protects assiduous labor involved in editing of materials	No protection of "sweat of the brow" (e.g., in careful proofing of a musical text resulting in minor changes to the content)
FRANCE: No right of fair use	Permits quotation of small portions of a protected work (score, recording, etc.) for academic and scientific purposes
GERMANY: Recording acceptable as primary instantiation of a musical work.	Primary instantiation of a musical work considered to be a score.

We have already held one symposium on the legal issues surrounding virtual scores (at Columbia University's Kernochan Center for Law and the Arts, May 2003) and two more are planned. One of these occurs in October of this year at the ISMIR (International Symposium on Music Information Retrieval) meeting in Barcelona, and another will follow in 2005 at Stanford University.

Methods of financial support for the making of such editions appear to play a substantial role in the current interpretation of rights and permitted uses of materials. The conventions that were in effect for the last half of the twentieth century can easily be viewed as having had strong cultural identities (to cite the theme of this congress). As these identities change, so do the organizational underpinnings of *Gesamtausgaben*. What seems to be needed now is a new cultural framework for such enterprises.

User-Interfaces

We have never regarded the development of user tools as our main goal, but we recognize their importance. We have sometimes been dismayed to see how weak has been the acceptance of computer technology by musicologists. This can be attributed to two phenomena:

(1) Students in the humanities are not encouraged to learn anything about how computers work. While this is fine by itself, some teaching of negative values discourages the curious and the energetic. Students are taught to prefer “off the shelf” applications with “glossy” graphical user-interfaces. Consequently, the possibilities for musicologists to investigate new kinds of applications and to refine and extend existing research methods remain invisible and many potentials unrealized.

(2) While musicologists are busy ignoring these opportunities hundreds of librarians, electronic publishers, computer scientists, software writers, mathematicians, cognitive scientists, audio engineers, and a host of other professionals are boosting the field of “music research” without the knowledge that the discipline of musicology exists and without the insights that musicologists could bring to their enterprises. The best students in our graduate seminars on music representation and music query come increasingly from these fields, that is, from outside musicology.

There are two paths to ameliorating this bifurcation between technology and musicology:

1. The first is to develop friendlier user-interfaces for software that is relevant to the tasks that musicologists are likely to perform. Some examples of efforts to do this can be found in the “JRing” browser of Andreas Kornstaedt (this symposium) and the “keyscapes” of Craig Stuart Sapp. Special skills are required for such implementations, but the best results will come, I believe, from collaborative projects involving musicologists, software designers, and perhaps engineers.

2. The other is to initiate collaborations between historical and systematic musicologists, for systematic musicologists may be best able to bridge the intellectual divide between technologists and “conventional” musicologists. In our own case, the most rigorous and constant of our data has come, somewhat surprisingly, from experimental psychologists. Music-psychologists bring rigorous statistical training together with a solid grounding in conventional repertoires in their search for better understanding how human beings relate to music. (In a sense they are practicing systematic musicology under another label.)

Towards the Future

The existence of growing corpora of electronic representations of musical works has re-kindled interest in systematic musicology in many interdisciplinary environments. From our perspective there is ample opportunity for cross-fertilization of historical and systematic musicology (both broadly defined) within this flurry of new activity.

Historical musicologists cannot maximize the value of their electronic editions, however carefully they may be edited, without using some of the new capabilities (e.g., the open-source music-analysis tools of the *Humdrum Toolkit*) that systematic musicologists have been using for more than ten years. Other tools of future value to musicologists include the intensive efforts of the present to develop music-repository-search capabilities which are cognitively informed (e.g., the recent work of D. Muellensiefen and K. Frieler) and music-analysis tools which give users access to multiple cognitive and perceptual levels of the work instantaneously (e.g., C. S. Sapp and A. Volk).

For historical musicologists, the hundreds of repertoires of the past which are unedited and unavailable for study or performance are also unavailable for analytical study. Modern tools for editing music requiring little enhancement to serve many other electronic purposes. Our hope is that the bridges that are to be built between the traditionally separate fields of historical and systematic musicology will enhance the value of musical data and the editions it is used to create not only for notation and source-control but also for sound and proof-hearing. It can also be extended for analysis and for methods of representing analytical results which improve their comprehensibility to non-specialists. This promises to open up new application areas in pedagogy and what might be called the “public outreach” of important repertoires. This is an ambitious agenda, but if we can achieve only a portion of it, it will significantly increase the value the arduous labor that is involved in creating editions and preserving our legacies for future generations.

References:

Kornstaedt:

Muellensiefen.

Sapp. (<http://ccrma.stanford.edu/~craig/keyscape/>)

Volk.

ESF

20-Aug-04