

Online Chopin Variorum Edition (OCVE)

Final report for the Andrew W. Mellon Foundation, November 2004

Contents

1 Introduction

1.1 Background and general overview

1.2 Project participants

1.3 Description of the project

- 1.3.1 Core issues and research questions
- 1.3.2 Project outcomes and deliverables
- 1.3.3 The research process
- 1.3.4 The consultation process

2 Results of the study

2.1 Musicological achievements

- 2.1.1 Summary
- 2.1.2 Excerpts from user reports
- 2.1.3 OCVE and other composers

2.2 Technical developments

- 2.2.1 Superimposition
- 2.2.2 Juxtaposition
- 2.2.3 Symbolic representation
 - 2.2.3.1 Extracting representations from musical scores
 - 2.2.3.2 Markup and metadata issues for symbolic representation
- 2.2.4 Annotation
- 2.2.5 Building the OCVE environment
- 2.2.6 Other technical issues
- 2.2.7 User trials

2.3 Content acquisition and library issues

- 2.3.1 Overview of library sources
- 2.3.2 Summary of outcomes
 - 2.3.2.1 Overall responsiveness
 - 2.3.2.2 Permission to display
 - 2.3.2.3 Image quality
 - 2.3.2.4 Pricing, delivery, and turnaround
- 2.3.3 Potential solutions for OCVE Phase 2

2.4 Licensing issues

3 Conclusions and recommendations for further work

4 Budget (project team and Mellon Foundation only)

4.1 Overview of expenditure

4.2 Budget statement (Excel attachment)

1 Introduction

1.1 Background and general overview

The Andrew W. Mellon Foundation awarded \$110,000 to Royal Holloway, University of London in December 2002 to fund an eighteen-month pilot study as the first stage within an ambitious and extended research project. The pilot began on May 1, 2003 and ended on October 31, 2004. Its principal aim has been to create an Online Chopin Variorum Edition (OCVE) in order to facilitate and enhance comparative analyses of disparate types of source material, attaining a level of manipulability far outstripping that manifested in extant printed editions of Chopin's music and indeed of any composer to date. The research has exploited emerging technical capacities for text/image comparison as well as recent musicological advances in cognate projects such as Chopin's First Editions Online (CFEO; see <http://www.rhul.ac.uk/Music/Research/ChopinsFirstEditionsOnline.html>), funded by the Arts and Humanities Research Board, and the *Annotated Catalogue of Chopin's First Editions* (Cambridge University Press, forthcoming 2005), itself the outgrowth of a four-year project funded by the Leverhulme Trust. Although its most immediate contribution is to the field of Chopin studies, the OCVE project is potentially relevant to the source materials of other composers' music from a range of chronological periods, and indeed to the very understanding of what a musical edition and a musical work might constitute. Longer-term goals include the assimilation of musical recordings within the Chopin variorum for the purpose of sound/text comparison, and the development of techniques allowing performers to construct their own performing versions from disparate sources. In addition to the musicological achievements to date, there have been important technological gains which themselves have justified the commitment of major financial and human resources, and which will ensure OCVE's immediate and continuing influence on research concepts and methodologies and on scholarly communications in general.

1.2 Project participants

OCVE has been directed by Professor John Rink (Royal Holloway, University of London) in conjunction with Dr. Marilyn Deegan and Mr. Harold Short (Centre for Computing in the Humanities, King's College London). It has employed the following personnel in addition:

- a Research Assistant, Dr. Christophe Grabowski (Royal Holloway), who provided considerable musicological input on a 0.25 part-time basis in months 1–12 of the project and *ad hoc* thereafter;
- a number of technical experts, notably John Bradley and Paul Vetch, based at CCH;
- other CCH staff, including Simon Tanner, Paul Spence, and Damien Doherty, who respectively handled the definition and monitoring of standards/specifications, web architecture, and web design, among other areas of responsibility;
- a part-time Administrative and Research Assistant, Dr. Danae Stefanou (Royal Holloway), who provided much greater project management support than had originally been anticipated, including technical input toward the end of the project; and
- a wide range of external consultants, engaged on an occasional basis to provide expertise on legal/copyright, technical, musicological, and other matters.

OCVE has also benefited from the support of administrative staff in the Department of Music at Royal Holloway, University of London, and at CCH.

1.3 Description of the project

1.3.1 Core issues and research questions

Although numerous variorum projects exist in the field of textual studies, many of which exploit the latest technologies with regard to image manipulation and collation/cross-referencing across discrete filiation chains, musicology has not previously exploited the application of such technologies to problematic source networks. The complexities

surrounding the Chopin source materials are particularly acute, more than justifying the development of an innovative variorum edition to help resolve or at least accommodate them. These arise partly from the diverse ways in which each work was drafted, prepared for publication, and subsequently revised in successive impressions, and also from the inadequate copyright protection available to Chopin in the 1830s and 1840s, which led to the multiple publication of most of his works in three different countries. As a result, a large number of discrete versions exist of almost all of Chopin's output, and each has a competing claim to represent his own intentions if not also those of his publishers. The primary sources comprise sketches, autograph or non-autograph manuscripts (whether intended for engraving or for presentation to associates), corrected and uncorrected proofsheets, first editions (usually released in three different countries and in successive impressions often bearing significant editorial revisions of disparate provenance), materials used by Chopin in teaching a number of students, and editions of pieces for which no other source material survives. All of these feature or will feature in OCVE either at this pilot stage or in subsequent phases of the research; for instance, seven different types of material have been included for each of the preludes in the variorum at present, in source pools comprising approximately twelve items each.

Although OCVE's primary scholarly goal is, as noted, to facilitate and enhance comparative analysis of disparate types of source material, we have also addressed the following core research questions:

- What is a musical "work," and how is the "work concept" that has prevailed since the mid-nineteenth century challenged by the Chopin sources?
- What is the best means of capturing in an edition the creative history implicit in the sources, ranging from the earliest sketches through to the last impressions of the first editions and beyond?
- How can the intellectual and logistical difficulties routinely experienced by editors in handling disparate source materials be overcome by means of technological support?
- In what ways might technology change the mode of presenting information previously contained within—or, conversely, uncontainable within—print editions? Moreover, how might technology fundamentally alter the musician's and the musicologist's understanding of individual sources, their often complex interrelationships, and their significance as artistic and cultural artifacts within a rich history of publication, pedagogy, and performance?

Our success in answering these questions can be gauged in terms of the project outcomes and deliverables, and also the scholarly impact that OCVE has had thus far and promises to have in future, as recognized in the user reports referred to below (section 2.1.2).

1.3.2 Project outcomes and deliverables

As the table below indicates, the pilot stage of OCVE has achieved all of the deliverables set out in the October 2002 proposal except for some which can be fully realized only in Phase 2 or beyond (e.g., investigation of the basis for editions of practical utility to performers). One other goal—the release of a print version of the current report—was modified in accordance with our intention to use the project website as the principal mode of dissemination (with a publicity flyer being distributed to increase general awareness) and in light of our wish to commission reports for online publication from a variety of musicological and technical experts. We also achieved important outcomes not envisaged at the time of the application—most notably the creation of a considerable amount of original scholarly material in the form of annotations and printable critical commentaries, with corresponding technical development to support the former.

Scholarly deliverables	
Goal as of October 2002	Current status
1) Analytical study of the complex issues involved in presenting and editing Chopin materials (and those of other composers) using conventional techniques.	<u>Achieved.</u> See John Rink, "The Final Score?," <i>BBC Music Magazine</i> (May 2004), 30–3; see also annotations within variorum, and "printable critical commentaries."
2) Case studies of Chopin's music as outlined under "Focus and timescales" in application.	<u>Achieved.</u> See online variorum edition.
3) Dissemination of the results of (1) and (2) above in one or more articles in high-profile musicological journals.	<u>Achieved.</u> See <i>BBC Music Magazine</i> article; cf. also annotations within variorum and "printable critical commentaries."
4) Investigation of the basis for editions of practical utility to performers as well as scholars, allowing combinations of elements from discrete sources in order to produce a composite version.	<u>Ongoing: for development in Phase 2.</u> See workshop reports at http://puffin.cch.kcl.ac.uk/ocve/content/activities.html ; see also user reports.
5) Exploration of the issues involved in the extension of the technique to editions outside the first-edition filiation chains.	<u>Ongoing: for development in Phase 2.</u> See user reports, especially those by Jim Samson and Nicholas Cook.
Technical deliverables	
Goal as of October 2002	Current status
1) Detailed report on the technical issues to be investigated during the research phase of the project.	<u>Achieved.</u> See section 2.2.
2) DTDs; architecture and design documents.	<u>Achieved.</u> See section 2.2 and online variorum edition.
3) Marked-up examples drawn from the witnesses investigated in the case studies, and presented in a web environment as per the architecture and design documents.	<u>Achieved.</u> See online variorum edition.
4) Prototype variorum edition on CD-ROM and the web produced as a result of the scholarly and technical work outlined above.	<u>Web version achieved.</u> It was decided not to disseminate the variorum edition on CD-ROM, in light of the more flexible and thus more appropriate web environment that had been developed. (See section 2.2.)
Managerial deliverables	
Goal as of October 2002	Current status
1) Reports from meetings of the Advisory Panel adumbrating the scholarly, technical, and managerial issues involved in planning the variorum edition, in accordance with the recommendations emerging from the proposed workshop/seminar as relevant.	<u>Achieved.</u> The Advisory Panel was extended to include all those who attended the workshops in London (June 2003 and October 2004) and Philadelphia (May 2004). Reports were received from most members; see also http://puffin.cch.kcl.ac.uk/ocve/content/activities.html .
2) Major report on the project and its achievements, published by the Office for Humanities Communication and widely disseminated throughout the musical and humanities-computing communities in print	<u>Achieved in part.</u> The current document is the "major report" referred to here. The team redeployed the funds designated for the OHC publication to commission musicological and technical

and online forms.	reports of potentially greater value to the team and other users more generally. High-quality flyer to be produced by CCH for distribution, to draw attention to these reports and to the variorum itself.
3) Project plan with costings and timescales for the full variorum edition project, prepared with reference to the feasibility study on copyright/permissions.	<u>To follow in February 2005.</u> See sections 2.3 and 2.4 regarding the feasibility study on copyright/permissions.
4) Preliminary investigations into how the online variorum edition will be published, distributed to, and maintained for the scholarly community.	<u>Achieved / Ongoing: for development in Phase 2.</u> The pilot's achievements to date in publishing and distributing the online variorum edition will be assessed and extended as relevant in Phase 2 and beyond.

1.3.3 The research process

Regular meetings of the project team were held in London throughout the project. The musicological investigations included identification of the various source materials relevant to the chosen case-study pieces (some of these materials existing in multiple states—e.g. in the case of printed editions—or in inaccessible private collections) and, eventually, close comparative study of these sources for the sake of the annotations and printable critical commentaries. Travel was required to a number of collections and for consultation between members of the team. Some additional material was acquired and technical tests on symbolic representation were undertaken as part of a scoping exercise for the Phase 2 research. The technical development at CCH is described in detail in section 2.2 below. Medium-quality reproductions of the sources were used at first in order to expedite the technical work, but a large number of digital scans of varying high-end specifications were ordered from a range of libraries in order to carry out technical feasibility tests and also to determine which libraries produce the best-quality material and on what basis (see section 2.3). We explored rights and access issues in a “Rights Seminar” in June 2004 held with the DIAMM team and involving various legal and library experts. Dissemination took place throughout the project in the form of workshop and conference papers, an article in *BBC Music Magazine* (see details above), and the emerging project website.

1.3.4 The consultation process

London, June 2003

The first of three OCVE workshops was held in June 2003 and involved musicological and technical teams setting an agenda for the OCVE pilot and for the pioneering work that will follow (see <http://puffin.cch.kcl.ac.uk/ocve/content/wshops/2003june.html>). One of the key recommendations was the inclusion of more scholarly content than had originally been envisaged, in the form of detailed commentary on the sources themselves, the philological significance of the variants revealed through the juxtaposition of sources, and the interpretive issues arising from those variants. As noted above, the project team incorporated this suggestion within the pilot itself, while other recommendations will be addressed further in Phase 2 and beyond, including the development of a workspace for the addition of personal and/or collaborative commentaries on the materials within OCVE; the creation of a more sophisticated indexing tool possibly linked to or drawn from the *Annotated Catalogue* (cf. also the CFEO project); the need to account for a range of musical layouts, idioms, and genres within the variorum technology; and the possible inclusion of historic and other recordings.

Philadelphia, May 2004

A second dedicated workshop took place in Philadelphia in May 2004 with musicological and technical experts from Britain, Canada, and the USA. The recommendations are listed at <http://puffin.cch.kcl.ac.uk/ocve/content/wshops/2004may.html>, including the need for:

- 1) a strong underlying symbolic representation of all the works to hold the information together and to better enable superimposition, juxtaposition, and recombination. With this representation we can create “polyglot” files that can be mixed and matched.
- 2) a search mechanism;
- 3) a facility to allow users to hear versions of audio performed;
- 4) strong structural metadata (e.g., via the Metadata Encoding and Transmission System);
- 5) reusable tools, including an annotation tool to be employed in many different environments/projects; “filters” so that annotations can be viewed in different “layers”; and a “paths” tool that can take users through preset examples, and that can also allow them to create their own paths;
- 6) attention to such issues as who will add annotations, and why; how performers would use electronic editions/scores; sustainability and longevity; superimposition as a “fringe benefit” rather than principal focal point; partial versus systematic encoding in working toward a more refined granularity and thus greater manipulability in terms of content rather than physical objects; and project planning in multiple stages and possibly involving multiple funding bodies.

General approbation was given to the notion of the “dynamic edition” under development in OCVE as against singular “Urtext” editions of dubious conceptual, aesthetic, and historical legitimacy.

London, October 2004

A final workshop was held in London in October 2004, involving the same musicological and technical teams as in June 2003 along with some new participants, and with both declaratory and investigative purposes. Demonstrations of the OCVE project website were given and breakout and plenary discussion groups held to assess progress to date and possible ways forward in subsequent phases of the project. The chief recommendations were summarized in individual user and technical reports received after the workshop, including those by Nicholas Cook, Jim Samson, and Cliff Eisen, as well as those from Benjamin Korstvedt, Michael Cooper, etc. (see <http://puffin.cch.kcl.ac.uk/ocve/content/reports/index.html>). These will be particularly useful in developing the Phase 2 proposal and at later stages in the research (note, for example, Jim Samson’s suggestion that “an extensive treatment of a small repertory should be the goal of this project,” embracing secondary materials such as later editions and sound recordings alike).

Other workshops and conferences

The OCVE team also participated in a DIAMM seminar in Oxford in November 2003, and the Rights Seminar in London in June 2004 (see section 2.4 below; see also the reports at <http://puffin.cch.kcl.ac.uk/diamm/content/news/workshops.html>).

The OCVE project was outlined by John Rink in a paper at a major conference in Melbourne in July 2004 (Symposium of the International Musicological Society—SIMS), in a session on “Music, Technology, and Research” organized by the Australian IAML and featuring a range of international participants. Professor Rink was also an invited participant in a panel discussion on historical musicological research and the new technologies (see <http://www.iamlaustr.org/sims2004.htm>). The project team presented a poster on OCVE and the related project, Chopin’s First Editions Online (CFEO), at the DRH Conference in Newcastle upon Tyne in September 2004. It should be noted that Professor Rink’s attendance at SIMS (partially funded by the British Academy as well as the OCVE budget) and the invitation of various international experts to the successive OCVE workshops were considered by the project team to be potentially more beneficial than a number of “solo” consultative visits to European and American scholars as had originally been envisaged, given the resultant opportunities for group discussion and dialogue. Further dissemination will take place at the Warsaw conference of IAML in July 2005, to which Professor Rink has been invited to speak on OCVE.

2 Results of the study

2.1 Musicological achievements

2.1.1 Summary

The above discussion indicates that OCVE's chief achievements include the reassessment of the "work concept" that has prevailed in music and musicology since the mid-nineteenth century; the development of a "dynamic edition" (as we refer to it) as against the static versions manifest in printed editions; and the creation of an innovative critical apparatus delivered by means of online technologies (e.g., pop-up annotation boxes) and structured in categories inviting novel modes of comparison between sources. These categories encompass changes made within the given source but at later compositional stages; apparent omissions in the source; differences in comparison to another source within the variorum edition; textual modifications of other kinds; inaccuracies/ambiguities; and detailed remarks on such matters as stem direction. Comparison with the critical commentaries found in standard scholarly editions reveals just how different the approach taken here has been.

The fact that users can reconstruct the creative history of Chopin's music, starting from the earliest surviving sketch material up to corrected reprints of the French, German, or English first editions brought out long after the original impressions were released, and thus in effect create their own "dynamic edition" of Chopin's music, means that OCVE counters any fixed conception of the musical work and resists the prioritization or privileging of one version over another. That does not mean that all sources are treated "neutrally"—on the contrary, as the critical apparatus indicates. Nevertheless, the aim is to allow informed choice on the part of the individual user, who generates his or her own "impression" of the music by tracing through its creative history and thereby gaining an unprecedented degree of understanding about how it changed over time. This process honours individual editorial prerogative even as it remains truer to the process of flux that characterizes all musical creation and recreation—to a much greater extent than in any print edition, thanks to the confluence of technology and musicology at the heart of the OCVE project.

2.1.2 Excerpts from user reports

It is worth summarizing some of the feedback received in our user reports, which can be seen in full at <http://puffin.cch.kcl.ac.uk/ocve/content/reports/index.html>.

- "... this is pioneering research that advances several light years beyond comparable attempts to present such materials online. The approach of the team has been to invest initial time and effort in methodology, deliberately restricting the repertory under investigation to two short preludes. It seems obvious that as the methodology becomes increasingly secure, the pace of the work will increase dramatically. This approach seems to me exactly the right one, and I am immensely impressed by what has been achieved so far." (Professor Jim Samson)
- "This is a fascinating snapshot of how digital technologies for music representation so far developed primarily for archival purposes can be adapted to create a musicologist's (or more generally musician's) workbench. As such it is piloting not only an invaluable resource for studying Chopin's music, but potentially a means of studying music in general—perhaps with application beyond music, too." (Professor Nicholas Cook)
- "This is one of the most exciting web-based music projects that I know of, not only for its specific content but also for the possibility that it can serve as a musicological and technical model for similar work on other composers. During the course of the meeting on 26 October 2004 I expressed my unreserved admiration for the project..." (Dr. Cliff Eisen)
- "The Online Chopin Variorum prototype site is highly promising. The site is very useable and loaded with excellent images of both manuscript and printed sources, well supported

by explanatory information. I found myself almost immediately becoming absorbed in studying the content of the site, which is sure testimony to its transparency and functionality. If the site did nothing more than bring together clear and precise digital images of the sources of Chopin's works it would be very helpful. Until now there would be no way to study these sources, many of which are located in widely separate physical locations, side by side. I found the quality of the images to be fully acceptable. The printed sources certainly come through very well. The manuscript sources may be more variable, but even the autograph sketch [A¹] of Op. 28 No. 4 is well-served, although the original document itself is less legible. Access to the images is very direct and well-presented, starting with the list of the available sources. The ease with which the different sources can be viewed is excellent. The option that displays an individual bar from all the sources is highly useful. I found myself treating this page as a home base, from which to migrate back and forth among the sources as I tracked down variants of specific textual details. The ability to move the image tiles around was helpful, as is the option of dragging one image on top of another to overlay them for comparison.” (Professor Benjamin Korstvedt)

- “This is an outstanding scholarly initiative possessed of extraordinarily potent implications for a wide variety of future users: performers/editors of Chopin’s music; scholars concerned with his compositional processes; performers and scholars concerned with editorial and other interpretive issues of nineteenth-century music; performers and scholars concerned with the music of other composers; performance-practice scholars; historians of print culture; and so on. Its potential for bringing together these generally (and unfortunately) disparate communities is strong. And the resourcefulness and technological proficiency of the technical team are exemplary.” (Professor Michael Cooper)

See section 3 regarding the suggestions for future development made within the user reports.

2.1.3 OCVE and other composers

It is worth noting that the participants at the various workshops organized by OCVE included experts on Machaut and other “early music” (Professor Daniel Leech-Wilkinson), Mozart (Dr. Cliff Eisen), Mendelssohn (Professor Michael Cooper), Bruckner (Professor Benjamin Korstvedt), Mahler (Dr. Paul Banks), various twentieth-century composers including popular musicians (Professor Nicholas Cook), and of course Chopin (Professors Jim Samson, Jeffrey Kallberg, and John Rink). This made it possible to hold valuable discussions of, for instance, the particularly thorny problems surrounding sketch material; larger-scale variants than those common in Chopin, as in the case of Bruckner; the limitations of screen size in terms of manipulating orchestral and operatic scores within a variorum edition; the possible need as a result of those limitations to separate scores into different instrumental parts or families of instruments (e.g., all string parts might be compared, subject to screen size issues); and other philological and notational issues. On the basis of these discussions and other consultation (e.g., at the SIMS conference referred to in section 1.3.4), the OCVE team has considered incorporating Chopin’s works for orchestra and various chamber ensembles either in Phase 2 of the research or at the first opportunity thereafter. This would allow both the technological and the musicological achievements of the pilot stage to be enhanced and extended—although the recommendation of Professor Samson to aim instead for “an extensive treatment of a small repertory” should also be borne in mind.

2.2 Technical developments

All of the technical developments for the OCVE pilot have been carried out at CCH. However, the various meetings and workshops held throughout the project with musicologists (especially those with ICT interests and expertise) were enormously helpful in alerting the project team to significant technical activities elsewhere in musicology, and in refining the

initial ideas about what might be possible or desirable in the present phase of research. For instance, the development of annotation tools was not included within our original work scheme, but the participants at the June 2003 workshop confirmed that they would be highly desirable within this pioneering variorum edition, given their strong desire for the provision of in-depth scholarly commentary. Similarly, we had not envisaged testing methods of producing underlying symbolic representations of the music, but the musicologists at the May 2004 workshop argued forcefully for this, and thus trials were carried out with promising results.

The variorum prototype is presented as part of the overall OCVE website at <http://puffin.cch.kcl.ac.uk/ocve/apps/index.html>. Originally we had intended to investigate both web and CD presentation, but the technical teams have been able to work within, and sometimes beyond, the constraints of current browser technologies. However, we have concerns about the limitations of those technologies and whether our requirements extend too far beyond them; these issues are explored below and in the technical report by John Bradley at <http://pigeon.cch.kcl.ac.uk/chopin/tech-report>.

The original project plan called for the investigation of three types of image manipulation:

- A **superimposition** of printed sources within defined/discrete filiation chains (in order to reveal variants)
- B **juxtaposition** of excerpts of materials (both manuscript and printed) from single or multiple filiation chains, viewable against chosen “base texts” (with sources for comparison to be chosen from icons) and in isolation (decoupage/montage)
- C **combination/interpolation** of elements from disparate sources in purposeful collations.

Developmental work on the third of these—combination/interpolation—was planned for a later stage (if deemed appropriate at that point), and while there has been much discussion of this, only limited technical work has been carried out to date, in the form of the tests of underlying symbolic representations discussed below. Considerable technical effort has been put into two other strands of development not originally envisaged: annotation and the creation of metadata to represent complex musical sources.

2.2.1 Superimposition

The superimposition of printed sources within defined/discrete filiation chains in order to reveal variants was an initial aim of the project, largely for the sake of efficiency gains in comparing large numbers of first-edition exemplars which might contain minuscule but potentially significant discrepancies. Certain other projects have investigated superimposition technologies for the comparison of musical sources, and initially we considered adapting Aruspix in particular, a technique developed in Switzerland (see <http://www.unige.ch/lettres/armus/music/devrech/aruspix/index.en.htm>). Contact with the Aruspix team revealed, however, that their techniques were not readily adaptable to the present project, so the technical team at CCH developed its own more suitable, highly innovative tool for the superimposition of sources. Described at <http://puffin.cch.kcl.ac.uk/ocve/apps/docs/index.html>, this allows users to move and overlay images within the browser for the sake of comparison of one-bar excerpts. When thus superimposed, the images become partially translucent, although moving the mouse pointer over one of the images makes it opaque. With careful positioning the user can make the image underneath opaque by rolling the mouse over any part of it not covered by the overlying image; this allows for close comparison of discrete passages—i.e., single bars. By alternating rapidly between opaque and translucent options, variants across apparently similar printed scores may be more easily discerned.

Discussion with the musicologists attending our meetings and a number of their user reports have suggested that superimposition in a purely image-based environment is interesting but of limited general utility, given the divergent nature of the sources visually and thus the difficulty of achieving overlay at a level greater than that of the individual bar. One user suggested that “The superimposition tool would be excellent for identifying discrepancies between different states of the same plates, but that seems too narrow an application to be really worthwhile.” This user suggested that a more profitable pursuit might be “to incorporate meaningful musical representations as well (e.g. with a reliable OMR routine linking the graphic images to a ‘base representation’). It would be possible for the system to generate ‘logical superimpositions’ (i.e., identify elements of variance between editions, graphically represented through different colours or whatever), which might be very powerful.” It has also been suggested by several musicologists that superimposition (and therefore detailed comparison) would best be effected using underlying symbolic representation. This remains to be done either in Phase 2 or, more likely, a later stage in the project, building upon the scoping exercises carried out to date and in the next several years.

2.2.2 Juxtaposition

The website has been designed such that users select a version from the list of all available sources for each work within the variorum edition, possibly with reference to the abbreviated or detailed philological information provided for each source. Clicking a link (not an icon, as originally envisaged) will display a full screen image of the chosen version, initially in “Page View” (where the entire source is shown) and then, by clicking on any given bar within the image, in “Bar View,” whereby all of the corresponding bars in the source pool for the work in question appear in a grid-like format, greatly facilitating the process of comparison that has traditionally made the act of preparing musical editions so difficult. Numerous functions ensure straightforward navigation—e.g., “Next bar” and “Previous bar” links, as well as a “Hide Annotations” option.

The most innovative feature is perhaps the user’s ability to **juxtapose** images at his or her discretion. Moving the mouse pointer over the dark grey rectangle above a bar image (containing the information about the version on display) changes the pointer and permits the dragging of images with the left mouse button. Users can arrange the images as they like, clicking on the “Reset Layout” button at the top of the screen to tidy the bar images back into a grid. When they have arranged the excerpts however they wish, users can then move between bars (using the “Next bar” and “Previous bar” links at the top of the page) but, remarkably, with **their personalized layout preserved for them**. This interactive feature of the OCVE website is certainly without precedent within musical scholarship and has few parallels outside it. (See section 2.2.5 for further discussion.)

Although the original intention had been to allow users to delimit larger excerpts than one bar at a time, this proved to be awkward for a number of reasons (not least issues relating to screen size and thus display capabilities), but we have not ruled out exploring this further in Phase 2 or beyond. The “Next/Previous bar” links and the ability to call up other sources with a single click allow proper musical contextualization to occur, which had of course been a primary purpose behind showing longer passages on screen.

2.2.3 Symbolic representation

As noted above, development of symbolic representations of the music being studied in the pilot was not an original aim of this present phase. However, all of the musicologists whom we have consulted have argued that we should aim to produce strong underlying symbolic representation of the works in our resource in order to hold the information together and to better enable superimposition, juxtaposition, and recombination. With representation of that sort, “polyglot” files that can be mixed and matched could be created (thus providing the raw materials for recombination). This would also allow a robust search mechanism to be

produced, together with other possible enhancements, such as a facility to allow users to hear audio excerpts of the music on display in front of them.

One of the challenges of this is to decide how elements from the different version would be identified and selected: it is unlikely that assembling a version merely by selecting whole bars at a time would be useful. For instance, the person compiling his or her own version might like to select distinctive accent marks or phrasing indications (slurs) appearing in one version but not in others, while other characteristics of the version with the unique accents or slurs might not be required. Thus, what is needed is an option to “take the accent marks throughout version X” rather than choosing entire bars only. To support this kind of selection by musical feature, it would be necessary not only to tag the music symbolically (so that each symbolic version would print out with the features contained in that version) but perhaps group within the markup conceptual musical features (sets of accents, sets of slurs etc.) across each version so that they could also be selected for purposeful combination.

It is still too early to say definitively, but for this sort of functionality the MEI symbolic scheme, with its potential to extend the markup for scholarly purposes, appears better suited than schemes like MusicXML which seem to be aimed primarily at symbolic representation for publishing. The task of deciding what features of this kind should be grouped together for selection in the MEI markup will require extensive work by the musicologists on the project if it is to be carried out.

2.2.3.1 Extracting representations from musical scores

Underlying representation is very difficult to achieve accurately, whether through manual or automated means, although the technical team has recommended a number of possible solutions:

- using automated techniques like Gamera (see the technical report by Ichiro Fujinaga at http://www.music.mcgill.ca/~ich/misc/OCVE_OMR/OCVE_OMR.html)
- rekeying, though it is very difficult to do this accurately, even with double rekeying (which is why music typesetting is expensive)
- making several passes through the material, with Gamera providing the first pass, with the proofing, correcting, and markup being done later.

As a part of the evaluation of symbolic representation the CCH team worked with Ichiro Fujinaga (McGill University, Canada), Craig Sapp (Stanford University), and Perry Roland (University of Virginia), who have extensive experience with music recognition software and music symbolic representation. We suggested that they try to generate symbolic representations of a complex page of Chopin’s music (from the Barcarolle Op. 60—one of the test pieces identified in our original proposal to the Mellon Foundation). An web-sized image of the chosen page can be found at <http://maple.cc.kcl.ac.uk/ocve/barcarolle-9.jpg>. Sapp used the OMR software SharpEye on an image taken from the University of Chicago’s online collection (<http://chopin.lib.uchicago.edu>). He then manually fixed the errors that were generated and added some features that SharpEye had not detected, although in the end not all the material on the original page was included. Overall, he reported that it took about 30 minutes for the one page. We loaded the resulting MusicXML into Finale and saved the page as a PDF file which can be viewed at <http://maple.cc.kcl.ac.uk/ocve/barcarolle-9.pdf> and which reveals the kind of information captured in this file.

2.2.3.2 Markup and metadata issues for symbolic representation

The derivation and display of underlying symbolic representations depends of course on the availability of systems to display the symbols in some exchangeable form and to add metadata where appropriate. A number of systems allow this, including the Guido Music Notation Format (<http://www.salieri.org/guido>), a formal language for score-level music representation; MusicXML, developed by Recordare (<http://www.recordare.org/xml.html>) to create an Internet-friendly method of publishing musical scores; and the Music Encoding Initiative (MEI), which has taken as a model the Text Encoding Initiative (TEI) which over the

last fifteen years has achieved a high degree of buy-in from the scholarly community for creating tools and standards for the representation of text. The MEI is primarily concerned with musical expression which is in, or can take, a written form, and is intended to provide some tools for the underlying representation of music. The MEI is based at the University of Virginia (<http://dl.lib.virginia.edu/bin/dtd/mei>).

Given that CCH has always supported and worked closely with the TEI, and also that the MEI design seemed to support the possibility of extensions to add scholarly functionality in the way that TEI is extendable, the decision was made to explore the MEI approach for encoding. Perry Roland, developer of MEI and the MEI tools at the University of Virginia, offered to carry out a trial in marking up a brief sample. His findings are reported at <http://dl.lib.virginia.edu/bin/dtd/mei>. Roland took the MusicXML file that Sapp had created with SharpEye and applied an XSLT transformation that he had developed to convert it into a partially complete MEI file. He then corrected some errors in this file and added features such as pedal and tempo markings that SharpEye had not recognized. He also produced a PDF file from his MEI representation using Mup software, the result of which can be seen at <http://dl.lib.virginia.edu/bin/dtd/mei/barcarolle.pdf>.

2.2.4 Annotation

As noted earlier, we did not originally propose having annotation as a component of the pilot. However, annotation is an interest of CCH (including support for annotations by users as well as resource creators) for a range of projects, and thus during the June 2003 meeting the technical team suggested it might be relevant to OCVE in view of the request by the musicologists at the workshop for detailed scholarly commentary. Their enthusiastic response led the technical team to experiment with building annotation tools—a non-trivial task in light of current browser technology. John Bradley's technical report (<http://pigeon.cch.kcl.ac.uk/chopin/tech-report>) describes the underlying model used for storing and displaying the annotations, which are added at the level of the bar, given that this is the smallest granule with which we are working. The user can choose whether to display annotations or not, both in "Page view" and in "Bar view," by clicking on the blue asterisk provided for each annotation. Providing an asterisk for individual annotations will probably be revised in future phases of the project, given that the score itself may be obscured by asterisks in bars with many annotations, to an extent that some users have found intrusive. Experimentation with mouse-overs, highlighting, and more discreet indications that a bar contains annotations is ongoing.

We are also investigating the most appropriate ways for users to add their own annotations, for both private use and in shared forums. Indeed, the annotation component in OCVE was originally designed to support users adding personal annotations as part of their research, not for the sake of publishing annotations by the musicologists attached to the project. For various reasons it transpired, however, that the annotation tools were used significantly only by OCVE project staff, and furthermore that the annotations that have been added in essence take the place of those within a printed critical edition (and thus potentially represent one of the project's major scholarly innovations) rather than being research notes made during the direct study of the OCVE materials. Although this scholarly commentary will be perpetuated and indeed extended at later stages (at the direct request of our musicological colleagues), we intend that future OCVE research will focus as much on how users might create and organize notes of their own, thereby supplementing those provided by the OCVE project team.

2.2.5 Building the OCVE environment

The technical report at <http://pigeon.cch.kcl.ac.uk/chopin/tech-report> describes the elements of the OCVE environment in some detail. Even working with a very small number of Chopin pieces for this pilot has demonstrated to the project team the huge complexity of the sources, given the number of different versions. Prelude Op. 28 No. 4, for instance, is presented here in thirteen different versions, and there are many more notated and performed versions that a full variorum might include; Op. 28 No. 20 exists here in twelve versions, and again much more material could be added. The sources themselves are internally complex, and need to be marked up into logical components that can then be juxtaposed and compared. An early decision for the purposes of the pilot was that our unit of study and presentation should be the individual bar, so the chosen pieces have been separated into bars; thus, all superimposition, juxtaposition, and annotation relates to the bar. Ideally, bar marking should be done in the high-quality scans, but for the purely pragmatic use to which the bar areas were to be put, locating them in the web-sized image was considered sufficient for the purposes of the prototype. Once this was decided it was relatively easy to locate software to allow the bars to be delineated; for example, software supporting the development of HTML imagemaps would do the job well enough, and to that end we used Mappedit. Imagemaps were not generated with actual URLs inside: instead, the recorded data was the bar number (or other bar-like object such as clef, time signature, and key signature).

One issue that arose during bar markup and annotation concerned scheduling and workflow. We began working on our chosen pieces using photocopies as that was what was available to us early on. In parallel with the start of the technical development, orders were sent to the holding libraries for the high-quality images that we wanted to present. Many libraries have considerable backlogs for digitization and photography, and thus we found ourselves having to use the photocopies for longer than we wished, given that the technical work had to move forward while waiting for images to arrive. The bar markup is currently a manual process that has to be done by someone who understands the musical notation (see below). This therefore cannot be easily transferred from the low-quality images to the higher-quality scans at the moment the work has to be redone. However, the technical team at CCH has been developing means of automating this process, using the mathematical “affine transformation” to provide a way to transform bar regions from one image to another of the same version (whether printed edition or manuscript) of the music. A further procedural point is that images were cropped and re-sized prior to uploading, with the intention of maintaining a level of consistency not only for superimposition but also for the sake of facilitated viewing. This is a somewhat controversial process in that some musicologists prefer “real-size” viewing; we will weigh this up during Phase 2, possibly building in multiple viewing modalities at that point.

It should be noted that bar markup is musically less straightforward than might be assumed, given what appear to be straight barlines in scores which seem to offer ready-made “dividing lines” ideal for the markup process. Such a view is inadequate, however, first of all because certain elements straddle barlines and may have to appear in the markup for each of the two bars in order to ensure comprehension of the material in question when either of them is viewed in isolation. Barlines in manuscript material typically are neither straight nor truly vertical, and of course a margin of space has to be seen on either side of a barline in order for a user to recognize it as such (as opposed, say, to the division between two bars occurring precisely on a barline, which would therefore appear in neither image and thus would leave the viewer in doubt as to whether the bar in question was complete, given that it would lack either a beginning or a concluding barline).

We have already stated that the development of the juxtaposition and annotation tools for this pilot has pushed current browser technology to its limits, and there has been a great deal of discussion about whether the complex editing functions needed for the variorum are indeed possible through browsers. This was an issue in two key areas of the development of the variorum environment:

- 1) the provision of the facility for users to move the juxtaposed images around the screen;
- 2) the complexity of the presentation of the annotations.

For (1), development has taken place in various research initiatives, among them those of Shipman et al. (1999) on the research support tool VIKI, and Kirschenbaum and Kumar (in progress) on the Lightbox software (<http://www.mith2.umd.edu/products/lightbox>). Both of these extend the basic browser model in rather different ways but to the same end—namely, to allow users to employ a two-dimensional metaphor to assist them in organizing their work with digital resources. In OCVE we tried to use the extended features contained in CSS, Javascript, and HTML to allow the user the kind of facility provided in Lightbox. Although the software functioned successfully on Microsoft's Internet Explorer versions 5.5 and 6 for Windows, it worked only in part on other browsers (and on the Macintosh), and on some contemporary browsers not at all. This seemed to be because our use of CSS and Javascript to create "moveable images," although within the documented range of these "standards," is beyond what most browser software writers had thought of, so our code eventually collided with various bugs and implemented limitations in certain browsers. Furthermore, this type of feature is outside the purview of what browsers are supposed to do, so it is unlikely that there will be much desire on the part of browser software writers to fix such problems. Indeed, any future version of Microsoft's browser might introduce new limitations that will prevent this from working.

The musicologists at our October 2004 meeting also observed some of the browser technology limitations, and suggested that a full-scale online variorum might ultimately need to provide scholars with the opportunity to download any software required for optimally viewing and handling scores, for example, in order to click and drag large sections. They also suggested that a CD-Rom version might be more functional in some ways than an online resource in itself, or that some kind of "hybrid" might be possible, while acknowledging that the current web platform was maximally accessible and thus particularly desirable. All of these points remain under review and consideration, and will be addressed further in the Phase 2 research.

In the case of (2) above, various problems with annotations can be identified. Although linked to a base document (in our case, images), annotations are in some senses independent. As an immediate way to handle this we chose to show annotations in a separate pop-up window. Unfortunately, the number of abuses for pop-ups has meant that they have become increasingly unwelcome in the browser community, and thus a growing number of browsers provide ways to turn them off.

Another key issue is that the creation of annotation text has the potential to stretch standard web browsers. Clearly, the only mechanism that browsers support to allow text to be put into an annotation is forms. In the context of HTML forms the challenge came from the fact that the text needed to support at least some display characteristics (paragraphs, bold emphasis, italics, etc.), and ideally it would show music symbols (flat signs, crescendo hairpins, and so on). In the OCVE prototype the project team has simply typed HTML tags into the "annotation create" form. There are no other widely supported mechanisms for this task that work across a range of browsers. The filing of annotations in the hierarchical structure available in OCVE also stretched the browser metaphor. The closest computer-graphical metaphor for this filing was that of moving files between folders, and the best metaphor for this requires support for drag-and-drop, which is not provided in any standard way in browsers. As noted, OCVE has a basic model for personal and public annotations (although at present these are not being made available to external users). Currently both personal and public annotations are stored on the server, but of course this is ultimately not the correct model for personal annotations. It should be possible to store personal annotations on a personal machine. Browser design, so much based on the client-server model, makes the implementation of this difficult.

2.2.6 Other technical issues

The OCVE web presentation employs two linked technologies. The general project information has been presented using CCH's semi-automated web publication suite, known as xMod, which allows for rapid, low-intervention generation of complete websites from TEI-based XML source data. Linked to this is the variorum edition itself, with the tools for manipulating the data.

2.2.7 User trials

Close consultation has occurred throughout the pilot with our target user group, and as is apparent throughout this report, their views have been fed into project developments at all stages. During October and November 2004, a set of expert users carried out intensive work with the variorum prototype: some of their comments have been reported in section 2.1.2 above, while the full reports are at <http://puffin.cch.kcl.ac.uk/ocve/content/reports/index.html>.

2.3 Content acquisition and library issues

2.3.1 Overview of library sources

Twenty-three institutions were identified as potential providers of source material. Together they hold fourteen versions of each of the two preludes in the pilot; unique versions obviously are held by a single library, while for other versions there is a significant amount of overlap. Where possible, it was decided to obtain multiple images of the latter from different libraries, in order to determine each institution's responsiveness, scanning standards, ability to conform to OCVE specifications, willingness to have images displayed online, pricing, and speed of delivery. This information will prove to be invaluable in Phase 2 and beyond.

A summary of holdings across libraries is presented below:

Source name	Prelude No.	Holding libraries	Total copies available across libraries
A¹ Autograph sketch, c.1839	4	PL-Wtfc	1
A² Autograph <i>Stichvorlage</i> , 1839	4, 20	PL-Wn	1
A³ Presentation manuscript, 1840; album of J.-M. DuBois de Beauchesne	20	F-Pn	1
A⁴ Presentation manuscript, 1845; album of Cheremetieff family	20	R-Mn	1
C^f Copy of A² prepared by Julian Fontana, 1839 (original lost; reproduction held by PL-Wtfc)	4, 20	PL-Wtfc	1
C^s Copy prepared by George Sand; album of G. Sand (private collection; reproduction held by PL-Wtfc)	4, 20	PL-Wtfc	1
F¹ French first edition, first impression, August 1839	4, 20	F-Pn	1
F² Corrected reprint of F¹ , late 1839	4	GB-Lbl, PL-Wtfc, US-Cah	3
G¹ German first edition, first impression, July 1839	4, 20	A-Wn, D-DI, D-Lem, D-Mbs, GB-Lam, GB-Ob, PL-KJ, PL-Tu, PL-Wam, PL-Wru, PL-Wtfc, US-Cah, US-NYp, US-Wc	14
G² Corrected reprint of G¹ , c.1868	4, 20	PL-Wekier, US-Wc	2
E¹ English first edition, first impression, early 1840	4, 20	D-Mhv, US-Cu	2
E² Corrected reprint of E¹ , c.1862	4	GB-Ob	1
E² Corrected reprint of E¹ , c.1868	20	IR-Dtc, GB-Lbl	2
S Stirling copy of F²	4, 20	F-Pn	1

J	Jędrzejewicz copy of F ²	4, 20	PL-Wtfc	1
D	Dubois copy of F ³ (i.e., later impression)	4, 20	F-Pn	1
Z	Zaleska-Rosengardt copy of F ^x (i.e., later impression)	4, 20	F-Ppo	1

2.3.2 Summary of outcomes

Twenty-five unique images of Op. 28 No. 4 and No. 20 were chosen for use in the variorum prototype (see the listings available at <http://puffin.cch.kcl.ac.uk/ocve/apps/home.jsp>). More than fifteen duplicates (predominantly of **G**¹) were received from five libraries; delivery of two further ones is pending. Three unique images have not yet been acquired: the Cheremetieff presentation manuscript of Op. 28 No. 20 (**A**⁵), and Zaleska-Rosengardt's exemplar of the French first edition of both preludes (**Z**). **A**⁵ is held in the Russian State Library in Moscow, which was unresponsive throughout the project, while **Z** is housed in the Bibliothèque Polonaise in Paris, which only recently reopened after extended renovation. Furthermore, images of three exemplars of the Barcarolle Op. 60 (one manuscript and two first editions) were acquired from the British Library in June 2004 and used for preliminary trials relevant to Phase 2 and with regard to the tests on symbolic representation referred to in section 2.2.3.

2.3.2.1 Overall responsiveness

The libraries' initial response times varied between 1 to 8 weeks, and most institutions replied promptly when orders were placed. Delays in other cases may have been caused by linguistic difficulties and also by their inability to meet the proposed scanning specifications. The most effective liaison took place with partner libraries in CFEO; this may bode well for the next round of ordering in OCVE Phase 2, thanks to the wider network of relationships that we have now established.

2.3.2.2 Permission to display

The first informal letter to libraries emphasized the non-commercial nature of OCVE, as well as the possibility of presenting images as lower-resolution jpegs within a restricted site to prevent unauthorised use of the high-quality digital images themselves. This satisfied more than half of the libraries that responded, but some institutions required a formal licensing application and up-front payment of a licensing fee for online display. Subsequent communications resulted in fee waivers from all but two institutions (i.e., all but one of those from which source material was acquired by OCVE). (See section 2.4.)

2.3.2.3 Image quality

Images of first editions were of a higher and more consistent quality than those of manuscripts, which often had to be derived from low-quality photographic or xerographic copies of the originals. A notable problem in this respect occurred with most of the material received from the Towarzystwo im. Fryderyka Chopin in Warsaw (PL-Wtfc); corrective action will be taken for future stages in OCVE. Colour variation was a frequent problem despite the inclusion of colour bars in the original tiff files; again, greater control mechanisms will need to be established in Phase 2 and beyond, drawing upon DIAMM's experience and possibly requiring on-site visits (even photographing) by members of the OCVE team.

2.3.2.4 Pricing, delivery, and turnaround

Prices per scan, including postage and service charges, ranged from £5 to c. £50 for a standard order and up to 50% more on an express basis. Three libraries provided images free of charge. Licensing fees were ultimately incurred only in the case of the Österreichische Nationalbibliothek, which has anomalously high prices and thus may not become one of our "preferred partners." Processing speeds primarily depended on the location of and backlog at each institution. Average turnaround times for European libraries

ranged from 1 to 4 weeks from the date of order; U.S. institutions were generally much slower, taking up to 10 weeks. This was partly the result of difficulties in transferring funds.

2.3.3 Potential solutions for OCVE Phase 2

Our experience to date suggests that ample time should be allowed for ordering from U.S. institutions as well as non-English-speaking libraries; communication with the latter could be enhanced by providing fully translated documents. Furthermore, image standards should be agreed systematically in writing and rigorously monitored within a predetermined workflow to provide timely feedback to any institutions that fail to meet them. We will establish strategic partnerships with participating institutions in order to facilitate communications and maximize mutual benefits. We will take an even more proactive approach to scanning, choosing to work only with those libraries who can meet our standards where we have the choice; we will define further benchmarking standards for image quality, particularly in terms of resolution, colour, and file type; and we will continue to conduct systematic and rigorous quality assessment procedures to monitor the standards of incoming scans. Finally, we will construct generous timeframes especially when dealing with U.S. institutions, taking further steps to avoid delays in delivery of images and, where possible, negotiating discounts and waivers especially (but not only) in the case of bulk orders.

2.4 Licensing issues

Both the nature of the sources involved in the OCVE pilot and the amount of third-party manipulation proposed in the form of the superimposition, juxtaposition, and collation/interpolation techniques raised a potentially complex array of copyright implications. As several versions of the two pieces within the pilot are manuscripts as opposed to printed editions, ownership is not always easy to identify, and a number of images are derived from sources whose originals are either privately owned or lost.

Licensing was therefore a major consideration from the earliest stage of our communication with libraries. None of the libraries that responded to our informal enquiries refused to grant permission to display images online, although in some cases a fee was initially proposed. It was consequently decided that the ordering process should be effected as soon as an informal agreement was in place with each library; formalized licensing documentation would then be sent to participating libraries for consultation at the end of the pilot project and in preparation for Phase 2.

It was decided that the project would develop two licence agreement documents with regard to copyright protection:

- 1) the "online display" licence, covering ownership and other aspects of the online digital images and aimed at the participating libraries;
- 2) the "website access" licence, covering access to the online images and aimed at OCVE's users.

The finalized drafts to be dispatched to participating libraries were adapted in part from similar licences used by DIAMM, with subsequent suggestions from the project team and from Professor Charles Oppenheim, who acted as an external copyright consultant. Furthermore, relevant issues were raised at the Rights Seminar on 11 June 2004, a report of which is available at <http://www.diamm.ac.uk/content/news/workshops.html>. Discussion during the seminar extended beyond the purposes of the present pilot for the sake of developing licences of greater longevity, although ongoing examination and consultation with libraries will be required during subsequent phases of the project.

Below is a summary of points covered by the current OCVE licences, followed by a list of considerations for the future, as discussed during the Rights Seminar.

Issues covered by current OCVE licensing

- The licence to display is non-exclusive.

- Copyright in the documents is vested in the holding library; in the case of manuscripts, the library should be authorized to grant the rights offered in the agreement.
- Copyright in the database collection of images displayed on the website is owned by the project.
- Collation and other scholarly work, including annotations, are permitted as long as they can be considered consistent with the non-commercial, academic character of this project.
- Users of the OCVE website will be able to access the variorum edition through a password, which they will receive upon returning their signed agreement by post.
- The project will try to prevent, as far as possible, unauthorized access to the website, and will ensure that all images are adequately represented. However, OCVE cannot be held responsible for the infringement of any third-party intellectual property rights.

General observations and issues for consideration in subsequent phases

- In the case of manuscripts, the project should take the initial risk and deal with any potential problems on an individual basis if they arise.
- The possibility of storing a user's work (such as annotations and analytical work) and communicating this to other scholars will be examined in Phase 2 of OCVE, and at that point the licensing implications thereof will be assessed. Since such work will be no more than an overlay, however, communication can be effected through links without the need to save and disseminate the actual image. Hence, this may prove to be a relatively straightforward process in terms of licensing.
- Project/user rights regarding interpolation and collation work will have to be thought out and re-examined according to the nature of future technical developments, if full copyright and ownership of the virtual work rest with the project.
- If OCVE eventually enables users to create their own edition drawn from source materials within the variorum, there should be no copyright problems since the work will be carried out on encoded content rather than physical images. However, in the case of manuscripts copyright problems may arise if owners are unwilling to share the rights of possession.
- If sound recordings are incorporated into subsequent phases of OCVE, originals rather than remastered versions should be used in order to avoid overcomplicated rights negotiations
- Both timescales and costings pertaining to OCVE's licensing needs should be addressed in planning for the next phase, including legal checks, transactions, translations etc.

3 Conclusions and recommendations for further work

This has been an overwhelmingly positive pilot project which has achieved far more than was originally envisaged. Our success has been possible thanks to the interest, commitment, and skill of the diverse participants. We have been very fortunate in the level of engagement from the musicological community as evidenced in the full and helpful user reports written by the musicologists and by the reports from our workshops. This endeavour clearly fulfills a need in musical scholarship generally, as well as in the study of the complex textual traditions surrounding Chopin's music. The development of a "dynamic edition" of this type is without precedent, and so too is the progress that we have made in challenging long-standing but problematic notions of the musical "work" and the singular Urtext edition valorized for generations by musicologists and musicians alike. The project's greatest contribution to musicology and to the humanities in general, however, is perhaps its symbiotic combination of scholarship and technology. In OCVE, technology not only facilitates the research process but shapes and determines it, fundamentally altering our understanding of the individual musical source both *in se* and within broad historical and cultural contexts changing over time.

The achievements of the technical team have been just as substantial, and they provide a firm foundation on which the project can confidently build in the next phases of the research. The tools that have already been scoped and built are highly innovative, and have been developed according to robust standards and using open-source building blocks. They are also sufficiently generic that they can be applied to the musical *oeuvres* of other composers, and indeed we will further develop some of them for use with a wider range of primary source data—for instance, the annotation tools. At CCH, this pilot has given the opportunity to test the combination of in-house generic web development tools (xMod) and the new tools described here.

In future phases of the project (and building on the suggestions of our user base), the following further work is possible:

- 1) in partnership with DIAMM, development of complex levels of permissions to allow personal, group, and public annotation work;
- 2) further development of the annotation tools, in particular creating hierarchies and taxonomies of annotation for sophisticated organization, with equally sophisticated navigational and search mechanisms;
- 3) research into the usefulness of affine transformations so that mappings of bars can be transferred from one version of a source to a closely related source, in order that bar separation and markup can proceed more swiftly;
- 4) further research into the use of OMR for creating underlying symbolic representations of all the sources;
- 5) further research into markup schemes for the underlying representations;
- 6) research into superimposition and collation of sources using symbolic representation instead of imaging techniques;
- 7) extension of the number and type of sources (including chamber and orchestral repertoire), and the number of witnesses of those sources;
- 8) the addition of recordings to the source pool;
- 9) research into the algorithmic separation of recorded music into bar structures so that recordings can be treated like the printed and manuscript sources for the purposes of comparison and collation;
- 10) further development of the juxtaposition tools taking account of user comments (including such features as the selection of multiple bars for comparison, ability to hide/show sources, facility to line up bars differently); and
- 11) development of person workspaces for scholars, where they could save collections of sources, annotations, personal notes etc.;
- 12) study of annotations and work practices by OCVE users (rather than OCVE's "publishers") to make sure the annotation tools support how users actually want to use the annotations.

All of these points, and no doubt others, will be more fully developed in the Phase 2 prospectus which we intend to deliver to the Andrew W. Mellon Foundation in February 2005.

4 Budget (project team and Mellon Foundation only)

Authors of report:

John Rink, Marilyn Deegan, Danae Stefanou