Schriften des Instituts für Dokumentologie und Editorik — Band 3

Kodikologie und Paläographie im digitalen Zeitalter 2

Codicology and Palaeography in the Digital Age 2

herausgegeben von | edited by

Franz Fischer, Christiane Fritze, Georg Vogeler

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Malte Rehbein, Patrick Sahle

2010

BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.d-nb.de/** abrufbar.

Online-Fassung
Herstellung und Verlag der Druckfassung: Books on Demand GmbH, Norderstedt 2010
ISBN: 978-3-8423-5032-8
Einbandgestaltung: Johanna Puhl, basierend auf dem Entwurf von Katharina Weber
Satz: Stefanie Mayer und LaTeX

Finding What You Need, and Knowing What You Can Find: Digital Tools for Palaeographers in Musicology and Beyond

Julia M. Craig-McFeely

Abstract

This chapter examines three projects that provide musicologists with a range of resources for managing and exploring their materials: DIAMM (Digital Image Archive of Medieval Music), CMME (Computerized Mensural Music Editing) and the software Gamera. Since 1998, DIAMM has been enhancing research of scholars worldwide by providing them with the best possible quality of digital images. In some cases these images are now the only access that scholars are permitted, since the original documents are lost or considered too fragile for further handling. For many sources, however, simply creating a very high-resolution image is not enough: sources are often damaged by age, misuse (usually Medieval 'vandalism'), or poor conservation. To deal with damaged materials the project has developed methods of digital restoration using mainstream commercial software, which has revealed lost data in a wide variety of sources. The project also uses light sources ranging from ultraviolet to infrared in order to obtain better readings of erasures or material lost by heat or water damage. The ethics of digital restoration are discussed, as well as the concerns of the document holders. CMME and a database of musical sources and editions, provides scholars with a tool for making fluid editions and diplomatic transcriptions: without the need for a single fixed visual form on a printed page, a computerized edition system can utilize one editor's transcription to create any number of visual forms and variant versions. Gamera, a toolkit for building document image recognition systems created by Ichiro Fujinaga is a broad recognition engine that grew out of music recognition, which can be adapted and developed to perform a number of tasks on both music and non-musical materials. Its application to several projects is discussed.

Zusammenfassung

Dieser Beitrag stellt drei Projekte vor, die der musikwissenschaftlichen Forschung bei der Erschließung ihres Quellenmaterials dienlich sind: DIAMM (Digital Image Archive of Medieval Music), CMME (Computerized Mensural Music Editing) und das Programm Gamera. DIAMM verbreitet seit 1998 digitale Abbildungen von Handschriften in

Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2. Hrsg. Franz Fischer, Christiane Fritze, Georg Vogeler, unter Mitarbeit von Bernhard Assmann, Malte Rehbein, Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 3. Norderstedt: Books on Demand, 2010. 307–339.

bestmöglicher Qualität und macht sie der Forschung auf diese Weise weltweit zugänglich. In manchen Fällen bieten diese Bilder mittlerweile den einzigen Zugriff auf die Quelle, weil Originale verloren gegangen oder stark beschädigt sind. Bei vielen Handschriften reicht eine reine Digitalisierung in hoher Auflösung allerdings nicht aus: Sie sind auf Grund ihres Alters, einer nicht sachgerechten Behandlung oder schlechter Lagerung beschädigt. Für den Umgang mit beschädigtem Material hat das Projekt mit handelsüblicher Software Methoden einer digitalen Restaurierung entwickelt, die verlorengegangene Daten vielfältiger Materialien wieder sichtbar machen konnten. Darüberhinaus verwendet das Projekt ultraviolette wie infrarote Lichtquellen, um eine bessere Lesbarkeit bei Rasuren, Hitze- oder Wasserschäden zu gewährleisten. Der Beitrag diskutiert sowohl die Bedingungen digitaler Restaurierung als auch die Interessen der bewahrenden Institutionen. Mit CMME und einer daran angeschlossenen Datenbank musikgeschichtlicher Quellen und Editionen steht den Forschenden ein Hilfsmittel bereit, um dynamische Editionen und diplomatische Transkriptionen anzufertigen: Im Gegensatz zur starren Form einer bedruckten Seite kann ein digitales Editionsprogramm Transkriptionen für eine Reihe von unterschiedlichen Auszügen und Versionen weiterverarbeiten. Gamera schliesslich ist ein Bausatz für eine automatisierte Bilderkennung, der auf der Basis einer Software für Musikerkennung entwickelt wurde und auf eine Reihe von musikwissenschaftlichen und anderen Quellenmaterial mit jeweils unterschiedlicher Zielsetzung angewendet werden kann. Anwendung und Nutzen in verschiedenen Forschungsprojekten werden diskutiert.

1. DIAMM

Although today many libraries have embraced the idea of creating online access to their collections, there are very few (if any) subject-specific collections that cross borders of collection and country. The Digital Image Archive of Medieval Music (DIAMM) is a medieval musicology project that has managed to do this. Established in 1998 to collect images of endangered fragments and sources that are difficult to access, it has grown to embrace all major sources of medieval polyphony worldwide, and is gradually creating images of all these sources and adding them to the online collection. DIAMM undertakes the rather unglamorous job of providing a mass of research materials to the research community (and increasingly to a much wider public) in a digestible and useful way.

At the time of writing some 3136 manuscript sources of medieval polyphony (i.e. excluding chant) dating from before 1550 are known to musicologists. The number changes frequently as new sources (usually only fragments) are notified at a rate of several a year, particularly as archives in Spain and Italy are better explored. Polyphonic music from the period approx 800–1550 survives in a number of beautifully preserved large, complete manuscripts ranging in extent from around 30 to 800 pages, and in size from about 12 x 15 cm up to books nearly a metre in height (designed so that a whole choir could sing from a single book simultaneously) from the Netherlands scriptorium of Petrus Alamire. Fragments add an enormous range to this corpus, equally varying in size from full sheets of huge choirbooks down to snippets the size of postage stamps, and ranging in extent up to large collections of as many as 50 leaves from larger books that have been dismembered but reunited more recently, either virtually or in the real world.

As with many medieval sources, there is an untapped wealth of fragments still to be discovered in the form of unremarked and uncatalogued endpapers, or boxes of bits and pieces that have not been thoroughly researched. Most existing library catalogues, prepared in the 19th or early 20th centuries, take no account of endpapers or binding fragments, and this neglect was reflected in the practices of rebinding well into the 20th century, when endpapers or paste-downs were routinely discarded when books were rebound. Some of the richest collections of fragments from bindings are to be found in libraries that did not enjoy a period of plenty in the Victorian era, when so many libraries embarked on wholesale rebinding of their collections.

The corpus is a realistic one to try to assemble under one umbrella: in comparison to other literatures the surviving witnesses are relatively few, but even with a controlled repertory the simple cost of creating images of all the sources and the time required to do so means that we are a long way short of completing the collection. Funding simply for creating and maintaining resources of this sort is difficult to come by, and the image collection presently grows only when a research project that requires images of a specific source is able to obtain funding to pay for digitization. We are also recipients of image donations from projects or institutions that have images but are unable (at present) to deliver them because of the cost of creating a delivery mechanism.

Central to the mission of DIAMM is to keep the resource free. Many online repositories of images are only available to those with personal or institutional subscriptions, cutting out the casual visitor, independent scholar, or interested amateur, all of whom have as much right to see the content as anyone else. Musicology is a small subject, and persuading a library with already limited resources to spend part of its budget on a subscription that might only benefit a very small percentage of its users is understandably difficult. Keeping access free has allowed users from all disciplines and from all walks of life to study, use, explore or simply enjoy this aspect of European cultural heritage.

Traditionally, research in this field has centred on the study of repertories preserved in large complete sources, often with a scholar dedicating their life's work to one particular manuscript. Access to other sources would depend on the scholar's ability to travel or to purchase photos or microfilms of inconsistent and often-dubious quality. Thus research in the field has tended to become the province only of senior academics, with their detailed research limited to the times when they could physically visit a manuscript.

When DIAMM began digitizing documents we had no model on which to base a projection of how the creation and online delivery of images of rare and valuable documents would affect the access to them in the real world. Our initial assumption was that fewer scholars would need to visit the documents, and thus digitization could serve a preservation function since the documents would not be consulted in person as much as they were before the images were available. In at least two cases, these digital images are the only form in which the manuscript may now be consulted. We were certainly correct that a huge percentage of the work that had previously required physical contact with the manuscripts no longer needed to be done on site. However the explosion in internet usage from the turn of the century meant that manuscripts that had remained in relative obscurity for decades (even centuries) were now being brought to much wider public notice, and the new interest this aroused meant that a far broader public were interested in them and therefore wished to visit them.

This was not necessarily bad: for many libraries and archives their funding relies on demonstrable public access to their collections. Website hit-counter or visit analysis engines do not really provide the right sort of data for user analysis: many website visits are accidental or speculative, and the type of user and usage is difficult to pinpoint. Physical bodies walking through doors is still a much easier way to demonstrate usage of a physical resource, so the traffic to archives arising from their online exposure was not unwelcome.

A further concern in making images available online was that the sales of printed facsimiles might suffer. Again this proved unfounded: the British Library found that when they made images available online of manuscripts that had been published in facsimile, sales of the facsimiles increase.

1.1. Changes in Scholarship

The first major change to scholarship that digital imaging introduced was that scholars no longer had to work in black and white because of the cost constraints of colour photography. Colour imaging costs no more than black and white in the digital medium (although there are storage costs, since colour images are typically about three times the size of grayscale ones). The second change is that the ability of projects such as DIAMM both to create high-quality images and deliver them online in a one-stop free resource means that scholarship is no longer limited to the insular study of a single manuscript (or group of manuscripts), nor to those who can afford to buy the surrogates necessary for a broad grasp of the repertory as a whole, but can now take on a far more holistic attitude and approach.

The idea that all manuscripts can be available either on one website, or somewhere on the internet so that they can be accessed from a workstation anywhere in the world will not be new to readers, nor will the rude truth that probably less than 5% of all manuscript holdings are available online, and a good number of these require a paid subscription to view them. At the time of writing DIAMM delivers less than 20% of the known sources of polyphony in our designated field (although that number is rising), the limitation being due firstly to the fact that not all libraries will allow their images to appear online (though a surprising number embrace the facility provided by DIAMM wholeheartedly, since it costs them nothing), and secondly to the simple cost of buying or creating the images to add to the resource. The project has developed an enviable reputation in quality imaging and adds monthly to the online content thanks to direct donation of images, outside funding for imaging of specific projects or collections, and collaborations with, and consultancy for, other projects in which images created for them by DIAMM are also delivered through the DIAMM website.

An unanticipated benefit of digital imaging at high resolution has been that in many cases the detail shown in these images surpasses what is discernible to the naked eye, meaning that whereas previously a surrogate (photo, microfilm etc) was a poor alternative to seeing the original, an excellent digital image could now be considered better than the original, and in the case of several important endangered manuscripts DIAMM images are considered by their owners good enough to warrant withdrawal of the original source from public access.

What this means for palaeographers is a vast new repertory of fine-grained detail that was never—or only rarely—available to study before. The examination on-screen of this level of detail has allowed scholars to see erasures that are not visible to the naked eye, to examine ink density that shows where a scribe lifted his pen and dipped, supplementing the evidence behind scribal concordance, to examine the direction of hair follicles on skin surfaces and their groupings to determine whether two leaves were cut from the same skin, and to highlight or adjust particular areas of a leaf digitally to restore or improve legibility.

1.2. Dealing with Colour

Unfortunately as exemplars of accurate colour reproduction digital images cannot be taken and studied at 'face value' in the same way as analog surrogates, particularly when that examination is only undertaken on a computer screen. If study of the image only requires the ability to read the text, then a high-resolution colour digital image will provide detail and focus (assuming it is properly taken) well beyond what a hard copy can provide, even in colour. However if colour recognition is crucial for research then for some, the potential of the digital image is far from being fully realized, while for others the inexpert use of digital images is misleading. With analog surrogates the control of colour and resolution is in the hands of the creator so that what is delivered to the end-user is controlled at the point of creation. In the digital medium, even if the image is carefully colour calibrated and produced at the highest resolution, the creator has no control over what happens to that image after it leaves their desktop.

Many computer users, particularly older users with deteriorating eyesight, set their screens to low resolution so that everything appears larger. Unfortunately this also means that everything on the screen is slightly fuzzy. Less common now (fortunately) is to find a monitor set to 256 colours, but it is sometimes deliberately done (particularly on older machines) to speed up screen redraw. This causes the loss of fine colour gradations, which break up into blocks. Screen colour is a long-standing problem: even matching screens calibrated with a spider and professional software often will not match in colour cast, and this level of calibration is virtually unknown outside imaging departments and professional print studios. It is also still surprisingly uncommon in digital imaging studios.

The digital facsimile therefore has not really come into its own—and possibly may never do so because of the huge variation in monitor type, quality and colour-cast, and the even greater variation in the needs and colour-sensitivity of users. The DIAMM website delivers images with a request to new users to perform a basic calibration on their screen, but we know that no matter how 'perfect' our images are at the point of capture, we can never deliver perfect colour to a researcher's desktop unless we deliver in print. It is interesting that even in the digital age, when books and images are very easily and cheaply available in digital form, the majority of people still prefer to carry a paperback novel, and still enjoy the look and feel of a beautifully reproduced facsimile accompanied by a scholarly study, preferring this to digital access.

Digital images cannot simply be used in the same way as conventional photographic surrogates. Some years ago a scholar wrote that the new technology would allow comparison of ink colours (among other things) leading to new discoveries about paper preparation, inks and scribal concordances. What that person did not understand was that colour is one of the most devastatingly misleading fields in the digital imaging world.

We can take our camera and lighting equipment all over the world and take pictures that we believe are 'the same' in quality and colour balancing, but without an understanding of colour profiles and gray balancing, lay users simply cannot compare any two pictures effectively. Most users simply cannot trust the colour that they see on their screens, nor can they effectively compare the colour of two images taken with different equipment or by different photographers even using the same equipment unless they know what they are looking for. Monitors come out of their boxes miscalibrated, with a variety of colour casts, so that even if the service delivering the images to the screen knows what it is doing in colour profiling, there is no guarantee what the scholar will see. Colour-blindness issues aside,[1] at least when we produce a colour facsimile we know that every person who uses it will be given material with the same colours. When we produce a digital facsimile or online image resource we have no control over what the user sees, we cannot even ensure they have their screen set to a sensible resolution or bit depth, so the images may appear blurry or in only 256 colours rather than millions.

The digital age that has been hailed as a great step forward for scholarship is still a minefield in imaging that is poorly understood, and one of the biggest pitfalls is the fact that those accessing the resources do not understand the potential deficiencies or shortcomings of the materials they are now using.[2]

Even assuming a 'perfect' digital image is delivered to the end user, there are problems in understanding that data. The problems however are massively compounded by image suppliers who create poor quality surrogates: images that were out of focus at the point of capture and have been 'sharpened'; under- or over-exposed images that have been level-adjusted; images of faded, damaged or dirty sources that have been 'improved' by the image manager so that they look clearer; batches of images that did not look the same as the manuscript when checked on an uncalibrated monitor, and have been colour-shifted by the operator to 'look right' on that monitor; imaging processes that apply lossy compression formats such as JPEG during the workflow, thus losing detail and colour refinement. In order to look at a digital image now scholars need to know that what they are seeing may not be an accurate representation of the original, and that accuracy may have been compromised in ways that they are not qualified to identify.

There are no easy answers to closing the gap between supply and understanding of the product by the end user. We can identify the disease, but we cannot cure it, nor can we ensure that image-suppliers operate to a baseline of quality that ensures the end-user is not being supplied with poor quality information, thus misleading them further: a huge number of major research institutions worldwide send images of spectacularly poor quality from their digital studios. Institutions employ outside suppliers to do their imaging work, and then rely on them for quality assessment of their own output. Several organizations (including DIAMM) have set down their own quality guidelines, most of which agree, but there is no protocol that ensures those guidelines will be followed, even if they are cited when ordering images.[3]

Deficiencies aside, the revolution in image-availability afforded by digital imaging coupled with delivery via the internet has immeasurably enhanced every aspect of

---

1 Roughly 10% of men are fully or partially colourblind. The condition is hereditary and sex-linked. Because the gene is carried by the X-chromosome vastly more men than women exhibit colourblindness. For more information see Bailey and Haddrill.

2 Cf Melissa Terras's chapter in this volume on "Artefacts and Errors: Acknowledging Issues of Representation in the Digital Imaging of Ancient Texts", 43–61.

3 DIAMM imaging-quality guidelines may be accessed here: <http://www.diamm.ac.uk/techinfo/quality.html>.

research that relies on contact with primary sources, and has provided us with surrogates that can be exactly duplicated (as far as we are able to determine) *ad infinitum*, without deterioration or degradation of the original or its copies. Unlike hard copies the digital original cannot become scratched or creased with use; it will not fade or change colour with age (again, at least as far as we are able to determine) and we can carry thousands of them around with us wherever we go in containers smaller than a paperback novel or, in the case of users of DIAMM, by accessing a web URL from any computer connected to the internet.

1.3. Dealing with Damaged Sources

Because the early years of work in DIAMM were primarily concerned with sources that were themselves damaged, one of the earliest exploitations of the images the project acquired was digital restoration. When DIAMM was conceived the idea of digital enhancement existed, but it had never been seriously explored or attempted with manuscript sources. There was no commercially-available software specifically designed for this sort of activity, and there still is not, since several software packages have all the necessary tools available along with all the other extras and gadgets designed for the artistic exploitation of digital image media. The decision was taken at the start of the project to use only existing established commercial software that would not involve the project in software development or maintenance. Adobe Photoshop was at the time the best package, and is still probably the most widely-used image-processing package.

The early fragments digitized by DIAMM had undergone a variety of distressing and damaging experiences that had left them in many cases unreadable, and in some cases barely recognizable as having ever shown music.

Most common in the variety of misfortunes to befall our fragments was their re-use as binding materials, either simply by re-using leaves as wrappers, or more destructively by cutting them up to make quire guards or using them as paste-downs or binding reinforcement. Many of the fragments were caked with glue and darkened by long contact with leather turnovers. Those used as wrappers were often so rubbed that no ink was apparent on the outer face at all. Whole manuscripts survived as distressing palimpsests, with only tantalizing glimpses of musical notation visible at the extreme edges of the over-written material, while others had vanished altogether, discarded during rebinding programs, and surviving only as offsets on the preserved boards to which they had originally been stuck down. That many of these fragments had been found and identified as music at all is a testament to the tenacity of one of the founding directors of the project, Andrew Wathey.

Faced with the publication of the fragments in a long-standing series of b/w facsimiles of medieval music sources, the directors decided that publishing an accurate representation of a nearly-black page was a pointless exercise: any detail that was apparent in colour would be lost in monochrome. The result was what might be described as a completely new discipline in palaeography: digital restoration. The idea gained currency as fast as digital imaging, but has grown without constraints. Many libraries gave us early pointers to the dangers of digital 'tampering' by expressing concerns that a 'good' restoration of an image might lead visitors to expect the original leaf to be in such a condition when seen in reality. Encountering the reality might then lay the owners open to accusations of poor husbandry, so in their eyes digital restoration had to be reconsidered as a form of misrepresentation rather than simply as an enhancement to research. This is not a factor particularly considered by those undertaking digital restoration for their own needs, but was certainly an issue for DIAMM when delivering edited images to our users of documents that belonged to other institutions.

1.4. Ethical Enhancement

The first step in enhancement is to decide at what point the process is simple adjustment, and at what point that crosses the line into editing. If an image is correctly captured—correct lighting and exposure; correct colour balance; sharply in focus at the finest detail level—and is therefore the truest reproduction possible of the original then there should be no need for post-processing adjustment to make it look more like the original. If the original source is unsatisfactory but the image of it is 'perfect', then *any* adjustment must be deemed to be editing.

With the reservations expressed by libraries in mind, I looked at ways in which to make a reproduction of a 'restored' original visibly obvious as an altered source. Captioning is necessary naturally, but a caption cannot express the extent to which the image may have been manipulated or adjusted. Quite simply the only way to demonstrate the extent to which an image has been 'improved' is to publish the two versions side-by-side in whatever medium they will appear. Any restoration appearing in print has the added problem of the authority that print confers: a published restoration may be seen as the only, or best, 'solution' to reading the document. The examination of these issues led to a re-appraisal of terminology: 'restoration' implies the return of something to an earlier state, and we could not say with any reliability that what I did to an image restored an earlier state of the source. The first step therefore was to refer to adjusted images as 'digitally enhanced'. My first efforts at digital enhancement concentrated on 'restoring' faded colours or darkening: so faded ink was darkened and accumulated dirt was lightened, increasing, or in some cases reversing the contrast of the original while retaining its basic colours.

The result was often something that could have been mistaken for the present state of the original document if seen in isolation. In some cases the change was so radical that a user might discard the true image of the original as a mistake of some sort. A better solution to this problem was to restore in bright colours: instead of selecting faded ink colour and darkening it, I selected the ink colour and replaced it with a completely unlikely colour such as green or purple. The pixels selected were the same, but there was no possibility that the end result could be mistaken for anything apart from an edited image. An unexpected benefit of this was that the contrast of the new colour against the old (green on brown instead of dark brown on pale brown) enhanced clarity and made reading the text much easier. The Stratford example (below) makes use of this technique.[4]

There are difficulties in some cases in deciding which items on a leaf are remnants of text. Enhancement requires this decision early on, and is where the first major mistake can be made that will compromise the results. The first step in restoration of a source like this should therefore always be a process that is applied to the whole image, without any selection of areas or colours. The simplest and least controversial method for this first process is a level adjust, where dark colours are made darker and light colours lighter. For many sources this is extraordinarily effective, and may be the only process required.

In the example given in Fig. 1 the centre gray and white sliders beneath the histogram have been moved to the left, lightening the pale highlights and the mid grays. The black slider has been moved to the right to darken the dark highlights—in this case the writing that we want to see.

Having enhanced the colours that are there, it is usually much easier to determine which belong to text, and these colours can then be selected and enhanced (usually darkened), although once again it is essential that first level enhancements of this type use selections that are applied to the whole image, to ensure that the electronic selection is not predetermined by what the editor hopes or expects to see.

The Stratford wrapper (Fig. 2a-b) is a very good example of how following this practice yielded results that were not those we expected. According to the cataloguing of this item, there is music inside the wrapper, but nothing on the outside. The first image of the back of the outside of the wrapper, when enlarged, revealed what appeared to be the shapes of musical notes in the top left corner, and what appeared to be a chain of ascending notes about a third of the way down in the middle of the page (just below the later writing). A basic level-adjust clarified that the notes on the top left corner were indeed music notation. Using these notes as colour selection reference points, the difference between the colours of the ink and of the background dirt, which were too similar to be separated by eye, was exaggerated, with surprising results. Most interesting was that the line of ascending note-heads in the middle of the page were

[4] Digital restoration techniques used by the author are described in the DIAMM Digital Restoration Workbook, available for download at <http://www.diamm.ac.uk/redist/pdf/RestorationWorkbook.pdf>.

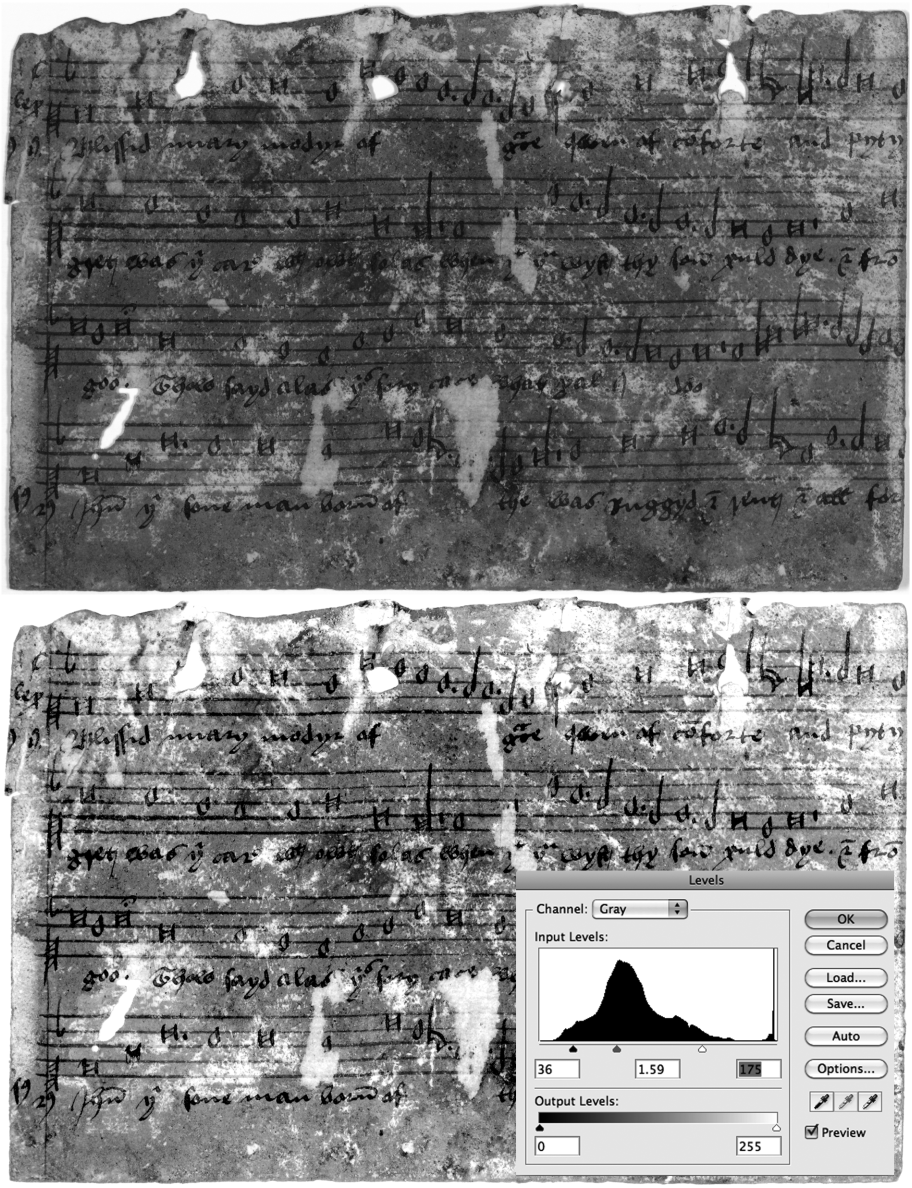

Figure 1. Cambridge, Pembroke College MS 314 supplementary envelope C, fol. 1v.

clearly not part of the text on the page itself, and may have been offset from a facing page, now lost.

Another sample of selection based on what was visible, rather than what we thought might have been there revealed a section of text (a second text to that underlaid beneath the music) that had gone unremarked even during the early stages of the restoration process (see Fig. 3).

The final example (Fig. 4) is of a leaf that was also used as a wrapper, but was not quite as damaged as the first sample. Various stages in the restoration process are shown, but it is difficult to decide at what point this leaf has been over-restored. Whole-image processes on the leaf revealed significant material, but although further processes improved some areas of the document, it disimproved others. The end user therefore needed a fluid version where some layers of processing could be turned on and off (easy in Photoshop, where the restoration is always done using 'layers'). In order to present a finished result that could be printed or shown outside Photoshop processes had to be applied only to selected areas of the document. The result is not a global adjustment process, but one that has had selective judgements applied, and thus is more of an editing process than a restoration.

1.5. Unethical Enhancement

I recently had cause to re-examine the ethics of digital enhancement when I was asked to photograph a document on which the owner was convinced certain words were written. I could not see these words, but the owner was determined they were there, and wanted to have the right sort of pictures, and to be taught how to digitally enhance them, so that he could prove what he passionately believed. The correct way to handle this situation would be to hand the picture to a third party who knew nothing about the content and ask them to see if they could find anything on it. The chances are though, that the third party would find nothing because they had no idea what they were looking for. At some point there has to be a decision about what type of enhancement to do, and that depends on what you may find. It helps if you know there is 'writing' there, but you don't know what it says: that way the writing can be made more visible by the disinterested party, and then it is up to the expert to read what it says and present the image to others for their opinion. This is not very different from the old-fashioned magnifying glass and palaeographer's trained eye: ultimately we have to trust the expert.

Enhancement that is undertaken to make an image show what cannot be revealed by 'justifiable' methods is no longer enhancement but editing. A colleague dubbed the results of this type of work a 'fake-simile'. Editing (or unethical process) is far from undesirable. Sometimes the results are sufficiently illuminating that the process is entirely justifiable, though how far to go in the pursuit of this process can present

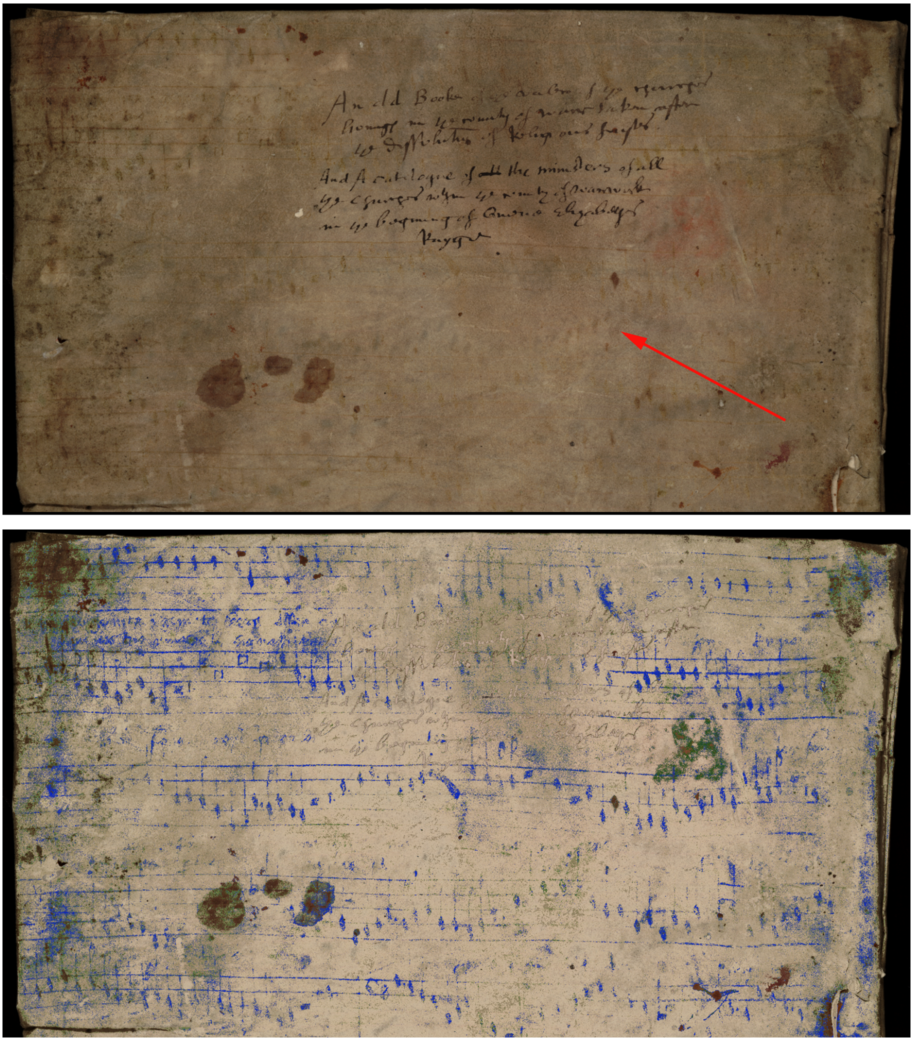

Figure 2. GB, Stratford-upon-Avon, Shakespeare Birthplace Trust, DR 37 Vol. 41, part of the back cover before and after restoration (suspect notes indicated by arrow).

Figure 3. GB, London, British Library Add. Ms. 41340 fol. 1 before and after restoration, © The British Library Board.

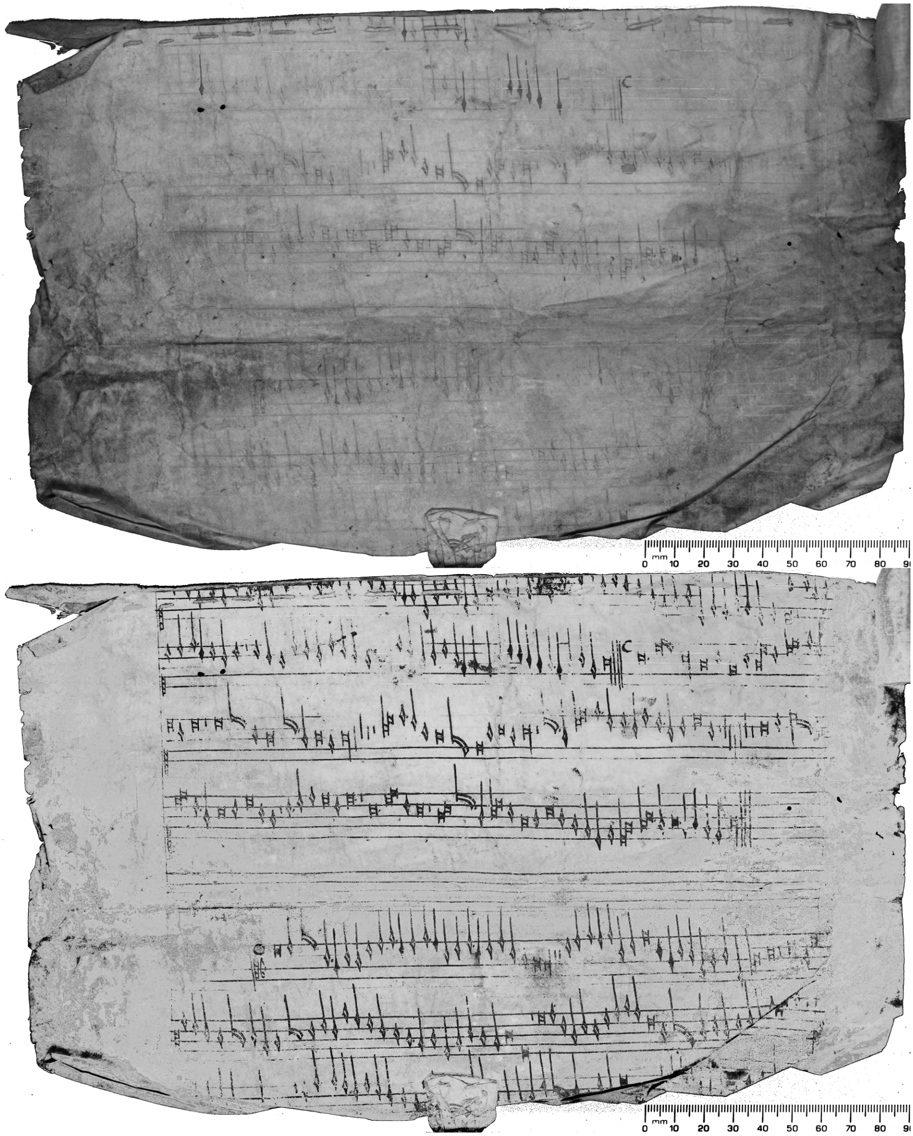

Figure 4. USA, Princeton MS 138.41 (Preece-Leverett fragment), photographed while in the private collection of Christopher de Hamel.

another stumbling block in presenting those results to a wider readership. As long as the restored version is always presented alongside as near a perfect representation of the original as possible, then the reader is still in a position to judge the veracity of the restoration.

The following example was taken from a paper manuscript damaged by ink burn-through. The first figure (Fig. 5) shows the original state of the manuscript. By magnifying the image on screen it was possible to see the difference between notes written on the reverse and those on the front face. Both notes have 'fuzzy' edges, but those on the front face have a more solid dark core. The duct of the script also gives clues to which notes belong on which face of the leaf. We can remove the notes that we are certain do not belong, and doing that removes considerable confusion from the visual field, making it possible to decide on a further group of notes that can be 'removed'. The process is repeated, also allowing for grammatical probability—so that where there is a choice between a next-door note and one at some pitch distance from the ones that we are sure of, the most sensible choice is the next-door note. Where there is heavy blurring, the position of the notehead can often be determined by using the stem, which—despite being hand-written—is usually of a uniform length. This process is best undertaken by someone who at least understands the vocabulary of the period, since that allows the restorer to determine which shapes are notes or rests, and which are other grammatical marks or nothing to do with the music. The reverse face can be restored simultaneously so that it can also be used as a reference for the repair of its partner.

At this late stage in the process working with the 'expert' musicologist in tandem with the restorer is necessary, since the expert knows the grammar of the period and the syntax of the musical work, and can make more informed decisions about which notes to remove and which to leave. However allowing the expert musicologist to do the work themselves from the start has proved inadvisable, since the temptation to see what they want to be there is often very hard to resist!

The second figure (Fig. 6) shows the 'repaired' version of the page. All but three notes on this page were restored with sufficient certainty that they were not in question. The three remaining notes were more difficult: by reconstructing the whole piece of music it was fairly clear what those obscured notes must be, but the ink that was apparent on the page did not match that expectation. At this point a heavy editorial hand was applied.

The process involved here is very different from that of selecting a colour and having the software find all instances of the same colour on the leaf. Since both foreground and show-through inks have the same values this technique of colour separation does not work. The process employed here is cloning—akin to simply rubbing out or deleting the colours that are not desired—where a clean part of the work of the same scribe with matching lines and line-spacing is copied over the area that we want to obscure. A leaf

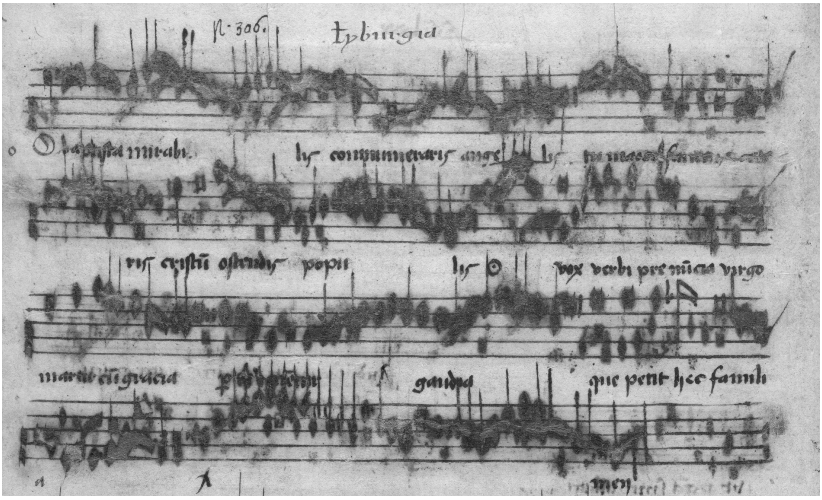

Figure 5. Italy, Museo internazionale e biblioteca della musica di Bologna, MS Q. 15 fol. 309v lines 1–4 as it appears in the MS. (reproduced by kind permission).

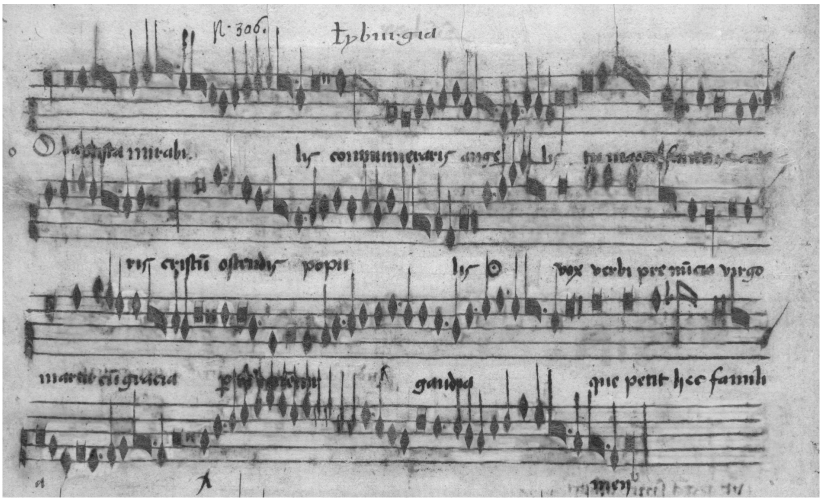

Figure 6. Italy, Museo internazionale e biblioteca della musica di Bologna, MS Q. 15 fol. 309v lines 1–4. Restored 'fake-simile' version, digitally altered by cloning from cleaner pages written by the same scribe.

Figure 7. MS Q.15 fol. 309v lines 1–4, cloned details only with base image removed.

from work by the same scribe that was undamaged (and cleaner areas of this page) was used to supply undamaged versions of notes, and clean sections of stave.

The final figure (Fig. 7) in this set shows only the cloned areas of the leaf, with the main leaf image removed. This shows what was overlaid on the main image to create the cleaned-up result. It was not particularly meticulously done since the object was readability, not perfection: a more elaborate restoration would eliminate the obvious points of changes in paper colour, and would clean up the remaining blurred notes that have been left here since work was not required simply to make them readable.

Cloning though is such a dubious activity that it cannot be called restoration (unless it is called 'unethical restoration'), since does not adjust information that is already there, it replaces it. In a moment of idle devilment while writing this article I inserted the word 'elephantes' into an unsuspecting line of biblical text from a medieval bible leaf (Fig. 8). It was not particularly carefully done and took about 5 minutes. It is included here only to demonstrate how easy it is to mislead ourselves and others when tampering with digital images.

The final example (Fig. 9) shows a process that is not so much unethical as simply not restoration, just the use of image-processing to elucidate what the eye and the skills of the expert musicologist can determine. The subject was a scan of a conventional b/w

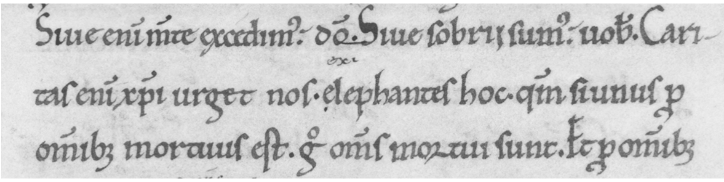

Figure 8. Digital fallacy: cloned text on a medieval bible leaf.

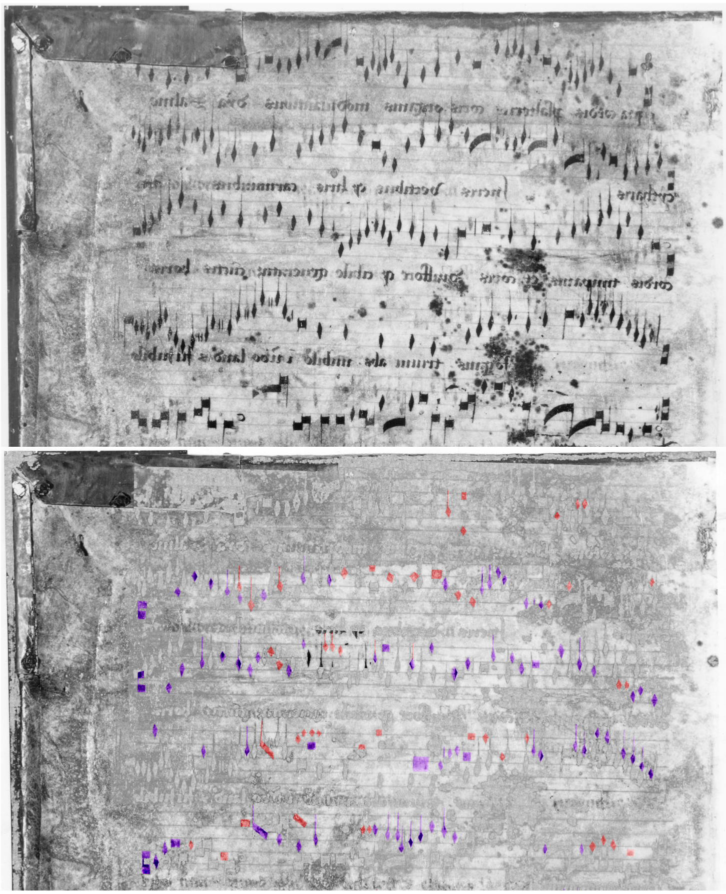

Figure 9. USA, Cambridge, Harvard University, Houghton Library, fMS Typ 122, fol. A, original and restored versions. (The first image has already been flipped.)

photo that was taken when a paste-down (still stuck onto a binding board) was treated with a chemical to make the parchment more transparent thus making the reverse of the page visible. In this case we want to ignore the notes on the superior surface, and enhance the ones on the reverse. The document is 'flipped' giving a mirror image, so that the notes we want to see are facing forwards, and the confusing notes that we want to ignore are reversed. Here again a trained eye is useful: the restorer in this case first 'removed' all the notes that he could see by duct and shape were on the front surface. He then flipped the document so that the reverse-face text was correctly orientated and enlarged the document on screen. Where he found shapes that he recognized as musical notation that he was sure did *not* originate on the front face he selected and then 'colorized' them purple. Notes that he was fairly sure belonged on the reverse (or that he considered were of dubious value) were coloured red, and it is evident from the second selection that there is some conflict between the two readings. The result was sufficient to identify the piece of music.

1.6. Delivering Metadata

The delivery of 'improved' images alongside damaged originals in the DIAMM content is a useful adjunct to the simple study of high-quality digital reproductions. However, key to examining and re-examining sources is easy access to existing secondary materials and metadata. As well as providing searchable electronic versions of the two seminal catalogues in medieval and early-modern musicology, DIAMM also provides page-images of the catalogues, allowing users to page through the books in their original typeset form. The cost and extent of the catalogues has, up until now, required those wishing to consult them to visit a research library:

CCM *Census Catalogue of Manuscript Sources of Polyphonic Music 1400–1550* (5 vols.);

RISM *Répertoire International des Sources Musicales: Manuscripts of Polyphonic Music c.1000 to 1550*, Series B IV (5 vols.).

There have been minor supplements to RISM Series IV, but essentially these two books are the only reference works that attempt to give a consistent level of information for every known manuscript of Medieval and Early Modern music in the world. 'Attempt' may be the right word, particularly in CCM: because the entries were compiled by local scholars, their descriptions of the documents tend to reflect their research interests, so one manuscript will have a long provenance history and no palaeographical information at all, while another will have extensive codicological information but no discussion of the musical contents or provenance. RISM is slightly more consistent, and provides invaluable inventories for all the manuscripts which the editors of CCM decided not to include (probably for reasons of cost).

At present the delivery of metadata through the DIAMM website is very limited but the full content of the back-end working database will be brought online during 2011 (possibly by the time this text is published). This will allow users to search on any or all of the following by using directed browsing as well as search forms:

- The full text source-description field;
- Text incipits;
- Full text transcriptions for many sources (in original and standardized spellings);
- Language database;
- Genre;
- Work title in original and standardized spelling;
- Voice part designations;
- Clefs and clef combinations;
- Scoring;
- Composer.

Inventories of each manuscript link to the images of each item together with editions of the works where those exist, and bibliographical links to articles relating to individual pieces. The manuscript metadata already links to bibliographical items (as does metadata for individual works within each source), but will also now link to a 'person' database giving scribes, dedicatee, binders, authors (of treatises), decoration style models, establishment patron, transcriber, owner etc.; manuscripts are also linked together in sets such as those manifest in the structure of the books (e.g. individual books in partbook sets, fragments of dismembered books) and other sets of a more intellectual type: links by copyist, illuminator, scriptorium, patron etc. will facilitate searches by or within groupings of manuscripts.

Musicology projects are increasingly working together so that work and resources used to create one database or delivery system are both available to other datasets and also transferable between datasets. There is a long-standing policy of sharing and collaboration between a number of large database-driven projects, and this will eventually become a considerable collaborative international resource in a way that would have been unthinkable only a few years ago. The participants: DIAMM, CMME, The Motet Database (which has been incorporated into the DIAMM database), Base Chanson (Ricercar) and Die Musik des Trecento (DMT), all deliver their data free to any user. Naturally cross-searchability of these datasets would be an enormous step forward, and is the goal of all the projects involved in a pilot phase of work to create a services model known as REMEDIUM.

The Motet database, Base Chanson and DMT are 'classic' metadata-driven resources, but CMME—Computerized Mensural Music Editing—had different origins for its dataset.

2. CMME

Theodor Dumitrescu began his work on creating an editor for early music notations while a Computer Science student at Princeton University, and continued to develop it later with research fellowships in Tours and Utrecht. He is currently a lecturer in the Music Faculty at the University of Utrecht, where development of more completed and polished versions of the software continues. The project is now able to deliver an accessible and intuitive version to researchers, as well as customize the software for specific projects (Dumitrescu; Selfridge-Field; Veit).

Mainstream music processing software is expensive, extremely complex and ultimately unsuitable for diplomatic transcription of medieval and early modern notations. Even if transcribing into modern notation this software is often too inflexible to allow the use of specialist editorial policies, and is also primarily designed for print output, not for online delivery. More important, these softwares are not XML-based, so input content is therefore not searchable. The online environment is particularly appropriate for the fluid presentation of materials and information when a single fixed presentation is undesirable. This is particularly true in a complex variorum environment that deals with more than a few primary sources although even comparing small numbers of sources print cannot be used to highlight different aspects of the variation between sources chosen by the user, only those that the editor has predetermined. Increasingly, and particularly with early music repertories, the needs of musicology require searchable content, and for the first time in music-processing this is now deliverable. Dumitrescu describes the purpose of his software thus:

> Without the need for a single fixed visual form on a printed page, a computerized edition system can utilize one editor’s transcription to create any number of visual forms and variant versions. The result is an entirely new form of critical music edition in which dynamically generated, user-configured formats remove the unwieldiness of multiple printed editions, replacing it with the concept of multiple states of a single edition. The early music editor’s task returns to the truly critical aspects of interpreting the text, rather than the ultimately unsatisfactory process of making presentation decisions which must limit the usability of the edition.
>
> (*Why online editions?*—CMME website)

CMME, as well as offering the facility to see music transcribed in a number of different editorial modes (chosen by the user/reader), can also be used to demonstrate variants between sources that can be chosen by the user, and highlighted with different visualisation techniques both on the score and in a configurable critical apparatus window (Dumitrescu and van Berchum).

As well as providing a clear WYSIWYG (What-You-See-Is-What-You-Get) interface for point-and-click editing using the same notation as in the original source (including red notation), a text editor, commentary tool, and support for notation features such as arbitrary proportions and colors, the viewer for finished editions includes an applet that allows the viewer to choose how they wish their score to appear. It is possible to view the music in separate parts as it appears in the original source, or as a score.

Having designed the software, Dumitrescu and his team are now utilizing it to create a set of dynamic electronic editions of important printed and manuscript sources. As well as serving as test materials for the software, this corpus represents a major contribution to the scholarly materials available in musicology:

> The music editions which populate the CMME corpus do not represent an anonymous mass of information punched in by disinterested data-entry workers; these are fresh editions, produced from primary source materials by musicological experts and kept at a high standard by our international board of editorial advisors. The reader of Shakespeare or Chaucer expects an edition which has been prepared with great care and knowledge acquired through study and experience, as does the performer of Lassus motets. There is no reason to waive these same requirements in online editorial endeavors.
>
> These principles of scholarly arbitration and peer review remain an important element in the formation of CMME "Editorial Projects." Corresponding in many ways to a volume in a printed edition series, an editorial project gathers up a collection of related compositions to be presented under the guidance of one editor or editorial group—for example, the contents of a single manuscript, the complete works of one composer, or a set of pieces known to come from a single court or city. As with any publication of early materials built on sound scholarly standards, a CMME edition provides the user with an introduction by the music's editors illuminating the historical, musical, and analytical context of the edition's contents. The web pages for individual compositions offer further commentary specific to each work.
>
> Beyond the set of editorial projects, however, users will quickly encounter a wider array of information in the Database section concerning musical sources, composers, and compositions which are not yet represented by music editions in the CMME. This network of contextual information [. . .] allows readers to explore the broader environment of the music editions in the corpus. The CMME meta-data collection limits itself largely to the one major element which is missing from the standard reference works on 15th- and 16th-century sources: the listing of actual contents, giving the names and composers of all compositions and their locations in the sources.
>
> (*Edition projects and "meta-data"—CMME website*)

Databases, however useful, are in a way an unexciting adjunct to codicology and palaeography: the concentration in current research projects on creating and populating them tends both to force a traditional (and necessarily) hands-on discipline to become distant and computerized, and further to imply that we no longer need to visit the original document, a fact that is manifestly untrue. Applying for funding simply to spend time with a source is nowadays almost unthinkable when building a case for a major research initiative. Both peer- and funding-pressure push us to create digital resources that we—and the technology available—may not be ready to create. We have only to look at the rapidity with which software has been created to 'read' historic newspapers and make a searchable archive from them to realize that almost anything we do now, with painstaking effort, may be automated in five years time. A very good example of this is Olive Software's *ActivePaper* which reads and classifies newspaper content, producing a very accurate searchable result.[5]

Many projects, but more significantly many funders, are seduced by the digital world, and the skills of manuscript handling, study and description are in danger. For many new students, if the information is not on the web then it doesn't exist, and that means that resources like DIAMM, for all the good they do, are also skewing study towards those documents that are accessible online, and away from those that are still difficult to access.

3. GAMERA

A case of research being driven by technology is to be found in the software creation of Ichiro Fujinaga, Assistant Professor in the department of Music Technology at McGill University, Montreal: Gamera. Gamera is not a packaged document recognition system, but a toolkit for building document image recognition systems. It makes the development of a new recognition system quite easy, though this still requires some time commitment. Gamera is a cross platform library for the Python programming language. Apart from providing a set of commonly needed functionality for document image analysis, Gamera additionally allows for custom extensions as C++ or Python Plugins and as Toolkits.[6]

I have often been surprised by listening to colleagues (particularly in classical and medieval subjects) wishing for a software that would recognize letter shapes in manuscript hands, when this is precisely what Gamera will do, and it has been around for some time and has been quite widely exploited. The software is teachable,

5 An example of the software in action can be found at the online archive of *The Scotsman* newspaper.

6 For a lengthy list of publications describing either the Gamera project itself or other research projects performed with the use of Gamera, see the list of publications on the Gamera website. Of particular interest to non-musicologists may be: Reddy and Crane; Canfield; Droettboom; Droettboom, MacMillan and Fujinaga.

so that when it finds characters it does not recognize it will suggest what it thinks they are, then when confirmed or corrected by the user it will re-evaluate the content of the document under consideration based on its revised database. Important here is its ability to *suggest* a reading, since sometimes a human reader will see a letterform and assume what it is from context. While this is often an effective way of dealing with damaged forms, or those obscured by interference of various types[7] it can mislead the reader into a completely incorrect interpretation of the material (Bowman, Tomlin and Worp).

The early development of Gamera arose from a desire to create a software that would read printed lute tablature. Tablatures are one of the most ancient forms of musical notation, showing players where to put their fingers, rather than the notes that will sound (which is what modern notation does). Lutenists still play from tablature, and some other tablatures are also occasionally still used.

Printed lute tablature has various features that make it both difficult and easy to deal with electronically: it is quite regular (the print was made from wooden type); the musical 'events' are evenly spaced, and there are a relatively small number of symbols in comparison with normal musical notation. In 'French' tablature six lines are used to represent the strings or courses on the lute. The top line represents the highest-sounding course. Letters are placed on the line to indicate which frets are to be stopped: the letter 'a' is the open string, 'b' the first fret, 'c' the second fret and so on.[8] Rhythm is indicated by a series of flags above the staff and, unusually for music in this period, the music is divided up into regular periods by barlines. The difficulty in teaching tablature to a machine is that because the letters printed on the lines are made with single-impression movable type, there are sometimes—but not always—gaps between one block and the next. The first task therefore was to remove the horizontal lines (deskewing the document along the way).

As hand-carved wooden type is prone to both natural irregularity and damage from use, the symbols are not very regular. In the first screenshot of the classifier below (Fig. 11; normally this appears in colour) the shapes in pale gray have been defined according to the information in the Gamera database. The shapes on a dark gray ground are those which the software is questioning. If the user clicks on a symbol its position is shown and highlighted in the document (in the bottom half of the window) so that it can be checked by context if it is not immediately obvious what it should be. Note that

---

7 E.g. when attempting to read text on the remains of wax tablets, where the wax has gone and all that is left is the indentation of sometimes several layers of writing on the tablet base, where intereference is caused not only by the layers of writing, but also by the grain of the wood; see Vindolana Tablets Online.

8 Italian tablature reverses the lines, so that the bottom line on the staff represents the highest-sounding course (thus matching the lines to the position of the courses), and the frets stopped are indicated with numbers (0–9) rather than letters. Obviously this system ran into trouble if courses higher than the 9th were required, and it extended into letters (x, etc.) from that point. Fortunately very high frets were rarely used in music of the period of the lute's heyday.

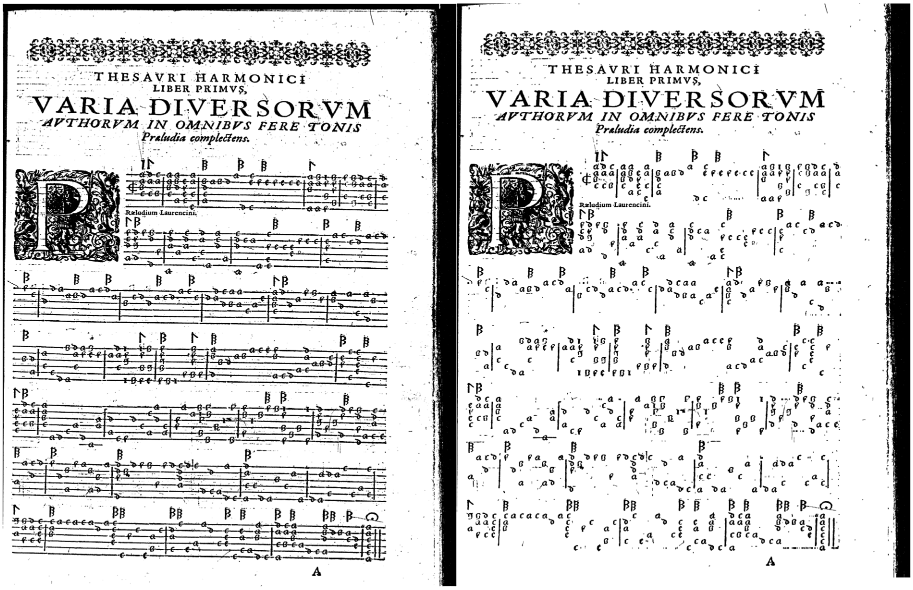

Figure 10. A page of printed lute tablature before and after staff-line removal.

although the symbol is being questioned, the software has nevertheless placed it in the correct category.

The user, having decided what the symbol is, assigns it to the classification list in the left part of the window, and the recognition database is updated accordingly.

The same process of line-removal can be applied to other musical notation such as neumes, commonly used to notate chant (Fig. 12).

The classifier database for that document set behaves in exactly the same way as for the tablature. The neume version is a significant step forward from the printed tablature, since this text is handwritten, and far more irregular than the tablature, since the scribe is compressing some figures to fit within a certain space. The spacing here is very irregular, but that does not cause a problem for the software. The neume shapes however are equally irregular—the first and second rows in the classifier show at least two different forms of each neume shape, but Gamera has successfully classified them, and only needs to have one confirmed.

The final example, from a handwritten Greek text was used to show that any shape system can be recognized, as long as the database has been given the basic set of symbols.

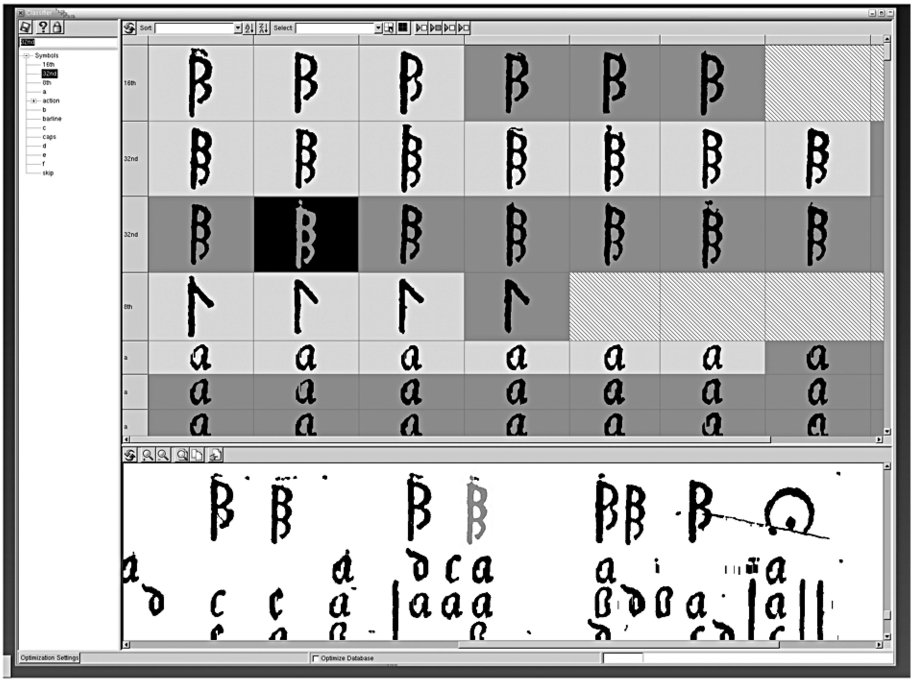

Figure 11. The Gamera classifier window, showing the identification of symbols on the pages shown in Fig. 10.

A custom version of Gamera was employed in the Online Chopin Variorum Edition (OCVE) where a massive number of digitized printed pages had to be marked up into individual bars. Although printed, there is very little regularity in the layout: in order to optimize page-space the engraver would compress the space between stave lines, and bars are never the same width. Using the horizontal line-removal feature, and then by creating a custom recognition system for vertical lines Gamera was used to automate markup of over 7000 pages of music, creating a list of co-ordinates for the bar lines that was then used to generate individual crops of each bar on each page, allowing a user to select one bar (or group of bars) and see the same bar (or set of bars) from the same piece from all the editions available in the dataset.

Fujinaga also prepared a custom version of Gamera for DIAMM, in which he input several hundred JPEG images of manuscripts in our collection, and programmed the software to examine the manuscripts and identify scribal concordances or instances

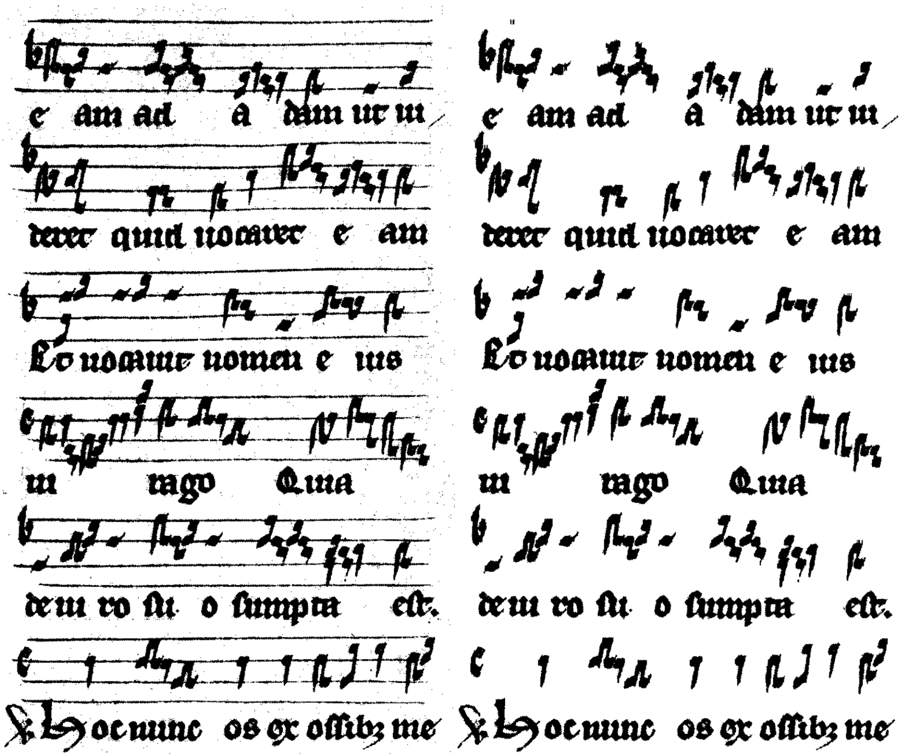

Figure 12. A page of handwritten neumes before and after staff-line removal.

where the scribe was similar. The software delivered a set of results that ranked the manuscripts by their similarity to the primary source expressed as a percentage (Fujinaga). Since it had scanned *every* source in our repertoire it had dealt with a number of materials that would have been extremely difficult for a human to handle, and the results it delivered were extremely convincing: it identified scribal concordances that we already knew about, and correctly ranked materials that were similar so that the expert palaeographer was directed to all the relevant documents in the collection without having to eyeball the entire collection themselves. Further refinement in the image collection to ensure accurate size reproduction for each sample would have enhanced the quality of the output further. Clearly the software has tremendous potential for exploitation in the fields both of script and shape recognition, and the types of study that

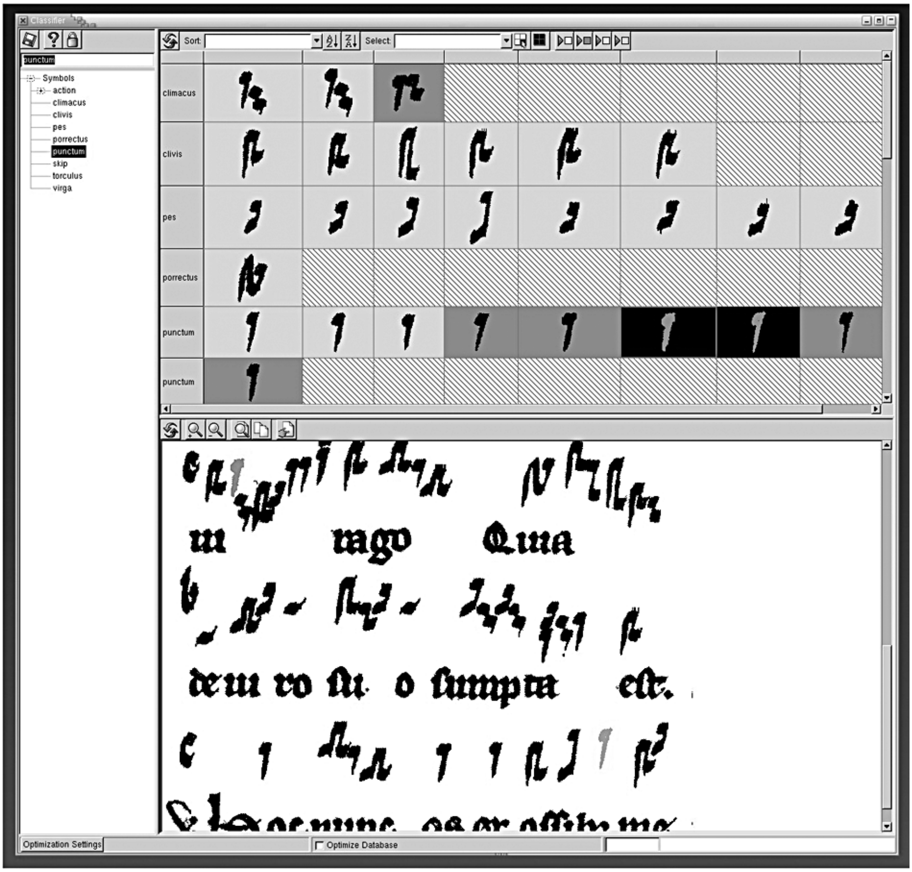

Figure 13. *Gamera* classifier window showing the identified shapes on the page in figure 12.

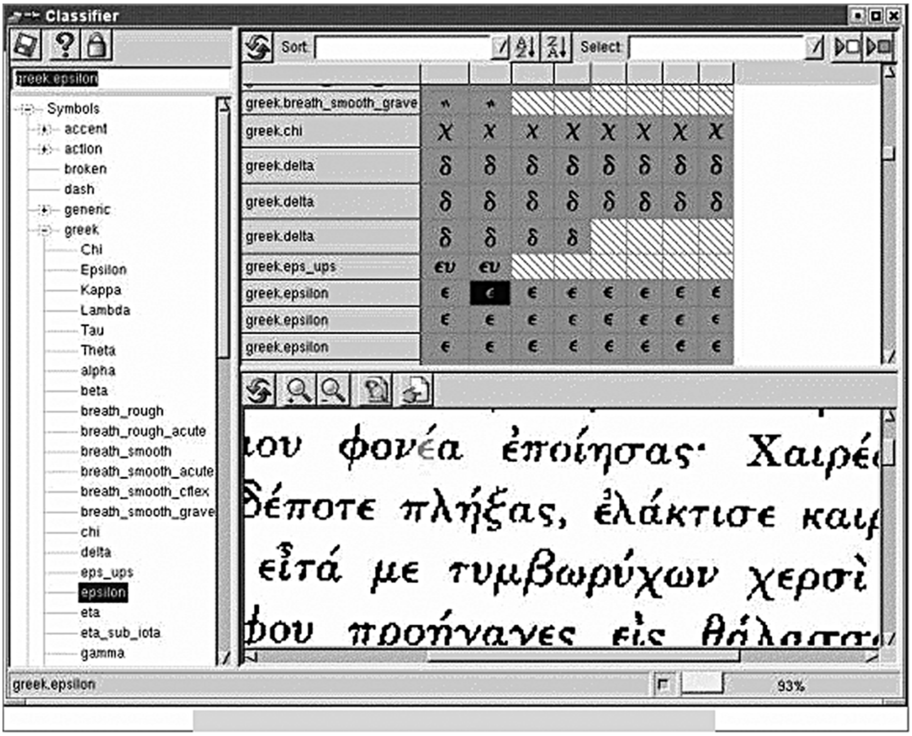

Figure 14. Gamera classifier window dealing with a page of Greek text.

can only be done by the analysis of huge bodies of data, most efficiently and accurately managed by computer.

4. Changes to the Discipline

Pre-digital palaeography has relied on the memory and the reportorial breadth of individuals: the visual memory of the distinctive characteristics of a scribal hand—and the memory of where that hand appears—is necessary to the identification of scribal concordances; in musicology an aural memory of entire musical works across a massive musical spectrum might be necessary to identify a newly-discovered fragment. Although a search can be narrowed down by using physical features such as notational idiosyncrasies, this can be misleading if the source was copied outside its original region or date period; in every discipline everything we read or see has to be retained if we are to make the best use of both primary and secondary, and we have to read or see everything available.

The limits not only of memory therefore, but of the amount of time we have to feed content into it, have shaped our research and study until now. There is a danger that reliance on digital and online content will cause researchers—particularly those new to the field—to limit their knowledge to what is available online, but that danger aside, we are now given the opportunity to search for the information we need across complete data sets instead of limited ones, and to search information that we might previously not have been able to find. In theory, a search for music manuscripts that might have been connected with the Electorate of Saxony can now be made reliably, with every reference delivered, without having to read the entire content of RISM or CCM and possibly miss a few references.

However, the fact that this is possible is dependent on the creation of the dataset and making it available for searching: research is still limited by the availability of content, and inputting that costs time and money, and funding is rarely available simply to create content. Dumitrescu's database of musical sources is searchable, but only the content that he has so far input can be searched. DIAMM will be eminently searchable, but the content is not yet complete, and data input still requires human analysis of, and decisions about, existing data. The creation of new information in the form of catalogue descriptions is even more difficult, since it requires a significant level of expertise, and most of those with that expertise do not have the time to devote to re-cataloguing work.

Fujinaga's software can be programmed to deliver a wide range of results as long as it is fed sufficient information, and its results carefully corrected to create a well-informed shape-database.

These are not the only sources of data available to aid research, but they give an idea of the type of resource that is now becoming, if not commonplace, at least an expected part of modern research. What remains is to populate these mechanisms with data, and that activity is fast moving from time-consuming and expensive manual input to automated systems. We have found that if we wait long enough, someone will find a way to automate what we have to do manually at the moment. The availability—and more significantly—searchability of data is inevitably going to change research, but although we can speculate now, only time will tell in precisely what manner.

Bibliography

Bailey, Gretchyn and Marilyn Haddrill. *Color Blindness.* San Diego, CA: All About Vision, 2009. <http://www.allaboutvision.com/conditions/colordeficiency.htm>.

Bowman, Alan K., Roger S. O. Tomlin and Klaas A. Worp. "Emptio Bovis Frisica: the 'Frisian Ox Sale' Reconsidered." *Journal of Roman Studies* 99 (2009): 156–174.

Canfield, Kip. "A Pilot Study for a Navajo Textbase." *Proceedings of The 17^{th} International Conference on Humanities Computing and Digital Scholarship* (ACH/ALLC), 2005. 28–30. Online: <http://gamera.informatik.hsnr.de/publications/canfield_navajo_05.pdf>.

CCM: *Census Catalogue of Manuscript Sources of Polyphonic Music 1400–1550* (5 vols.). American Institute of Musicology. Suttgart: Hänssler-Verlag, 1979.

CMME: *Computerized Mensural Music Editing.* <http://www.cmme.org/>.

Craig McFeely, Julia. "Digital Image Archive of Medieval Music: The evolution of a digital resource." *Digital Medievalist* 3 (2008). <http://www.digitalmedievalist.org/journal/3/mcfeely/>.

DIAMM: *Digital Image Archive of Medieval Music.* Oxford: University of Oxford, 2008. <http://www.diamm.ac.uk/>.

DMT: *Die Musik des Trecento.* Hamburg: Musikwissenschaftliches Institut der Universiät Hamburg, 2001–2007. <http://www.trecento.uni-hamburg.de/>.

Droettboom, Michael. "Correcting broken characters in the recognition of historical printed documents." *Joint Conference on Digital Libraries* (JCDL'03) 2003. 364–366. Online: <http://gamera.informatik.hsnr.de/publications/droettboom_broken_03.pdf>.

Droettboom, Michael, Karl MacMillan and Ichiro Fujinaga."The Gamera framework for building custom recognition systems." *Symposium on Document Image Understanding Technologies* 2003. 275–286. Online: <http://gamera.informatik.hsnr.de/publications/droettboom_gamera_03.pdf>.

Dumitrescu, Theodor. "Corpus Mensurabilis Musice Electronicum: Toward a Flexible Electronic Representation of Music in Mensural Notation." *The Virtual Score: Representation, Retrieval, Restoration*, ed. Walter B. Hewlett and Eleanor Selfridge-Field, *Computing in Musicology* 12. Cambridge, MA: MIT Press, 2001, pp. 3–18.

Dumitrescu, Theodor and Marnix van Berchum. "The CMME Occo Codex Edition: Variants and Versions in Encoding and Interface." *Digitale Edition zwischen Experiment und Standardisierung: Musik—Text—Codierung.* Ed. Peter Stadler and Joachim Veit. Beihefte zu Editio. Tübingen: Niemeyer, 2009. 113–130.

Fujinaga, Ichiro. "Creation of a Searchable Database for note shapes." *DIAMM. Technical Content and Web Delivery Workshop. 29–30 July 2004. St Hilda's College, Oxford.* <http://www.diamm.ac.uk/redist/pdf/July.pdf>. 21–25.

Gamera Project. Krefeld and Mönchengladbach: Niederrhein University of Applied Sciences, 2008–2010.<http://gamera.informatik.hsnr.de/>.

OCVE: *Online Chopin Variorum Edition.* London: King's College London, 2010. <http://www.ocve.org.uk/>.

Olive ActivePaper Archive. Aurora, CO: Olive Software, Inc., 2010. <http://www.olivesoftware.com/

Reddy, Sravana, and Gregory Crane. "A Document Recognition System for Early Modern Latin." *Chicago Colloquium on Digital Humanities and Computer Science: What Do You Do With A Million Books (DHCS 2006).* Chicago, IL: University of Chicago, 2006. Online: <http://dhcs2006.uchicago.edu/abstracts/reddy.pdf>.

Ricercar. Programme de recherche en musicologie. [Base chanson.] Tours: Centre d'Études Supérieures de la Renaissance de Tours (UMR 6576 du CNRS). <http://ricercar.cesr.univ-tours.fr/>.

RISM: *Répertoire International des Sources Musicales; Manuscripts of Polyphonic Music c.1000 to 1550.* Series B IV (5 vols.). Munich: Henle-Verlag, 1965 ff.

The Scotsman. Digital Archive. Edinburgh: The Scotsman Publications Ltd., London, 2010.

Selfridge-Field, Eleanor. "XML Applications in Music Scholarship." *Music Analysis East and West.* Ed. Walter B. Hewlett and Eleanor Selfridge-Field. *Computing in Musicology* 14. Cambridge, MA: MIT Press, 2006. 21–40.

Veit, Joachim. "Musikwissenschaft und Computerphilologie—eine schwierige Liaison?" *Jahrbuch für Computerphilologie* 7 (2005): 67–92. Online: <http://computerphilologie.tu-darmstadt.de/jg05/veit.html>.

Vindolana Tablets Online. Oxford: Oxford University 2003. <http://vindolanda.csad.ox.ac.uk/>.