A topography and taxonomy of digital musicology

Eleanor Selfridge-Field, Braun Music Center #130, 541 Lasuen Mall, Stanford University, Stanford, CA 94305-3076, USA. esfield@stanford.edu

Abstract

How does one make sense of the rapidly growing array of resources for musicologists online? In part, the answer depends on what specific interests are topical in the local environment. User needs vary widely. This contribution examines some organization issues in large repositories that available without cost in Europe and North America. Obstacles to finding sites with relevant material are many and generally parallel those familiar from using physical libraries.

Summary

The access for rare materials of interest to musicologists has grown exponentially over the past decade. Yet every collection has its own character and it can be daunting for users to find their way through the thicket. The character of a digital collection usually reflects the emphasis within the institution. Conservatories, for example, can be counted upon to hold large collections of scores and parts. University and national libraries, on the other hand, have multi-faceted collections that may include early liturgical books that include chant notation; lithographic collections that include images of singers and conductors; book collections that include libretti; sound archives; historical map archives; and much else that, although appearing to be peripheral to music, can be valuable in understanding cultures of past times. In a limited amount of space, we can describe only a few highlights of some of the largest, most easily accessible collections. For a fuller picture, we invited readers to peruse our website Digital Resources for Musicology (http://drm.ccarh.org).

Article

When in 2015 we began to assemble a reference list of open-source digital projects (Digital Resources for Musicology, or DRM), we did not appreciate how diverse the organization of digital collections could be.[1] Our immediate motivation was to assure continuing access to previously discovered cites, which, we noticed, might be absent from a saved bookmark the next time we wished to find it.[2] The present short commentary aims to give some sense of the terrain we have covered in our search for viable resources that are available to anyone and how to make good use of it in the immediate future. Digital musicology stands at a very early point in its development. Trying to forecast its long-term hits and misses would be premature. This is especially difficult for a nation such as Croatia, where funding for digital projects in the humanities is currently at or close to zero. Its disadvantage is increased by widespread interest in recent repertories, which are almost universally controlled by copyrights that limit the likelihood of web availability.

Here in the US most users would consult Google to find pertinent resources. Yet it can be difficult to locate academic repositories online with Google.[3] Some universities maintain lists of digital resources, particularly for finding licensed-access materials (those for which institutions must pay an annual fee). In smaller libraries, access to licensed collections—MGG-Online, Oxford Music Online (i.e. Grove), RILM, RIPM, et al.—may be unaffordable. Countries in Eastern and Southern Europe are among the disenfranchised. This divide in access is somewhat less evident in establishing digital repositories. Here we might hope that EU funds can enable more digitization projects in Croatia and neighboring countries.

---

[1] The editorial "we" refers to the Center for Computer Assisted Research in the Humanities (http://www.ccarh.org), an affiliate of the Packard Humanities Institute, located at Stanford University.

[2] Links can fail for several reasons: the site is not maintained, the url (web address) is changed, the website is restructured, or the focus of the overall project shifts with concomitant effects on content.

[3] Eleanor SELFRIDGE-FIELD, *Il Maeftro di Mufica* [sic] or Quality Control in the Virtual Library, *Fontes Artis Musicae* 62 (2015), 62-77, investigates the reliability of snippets for Google Books. It reviews entries in several languages (English, French, German, Italian, Spanish, and Dutch) and finds that all cognates for the word music are prone to be misreported as mufik (f instead of s) a small percentage (3%) of the time. This problem also affects a number of other words across the same languages, with German texts suffering the greatest likelihood of semantic loss. Two other inscrutable traps for unsupervised scanning are that Google's process fails to observe changes in fonts that signify changes in narrator or textual function (i.e. picture captions) as well as lines separating running text demarcating the start of footnotes. These lapses have consequences ranging for hilarious to catastrophic. In short, Google searches are not always as reliable as we take them to be. Together with Google's privileging of commercial material, academic websites descend in the limbo of lower rankings.

One consequence of funding disparities is that smaller countries around the globe may be represented within initiatives elsewhere. *Renaissance Music of Croatia*, for example, is one component of the Ricercar set of projects at the University of Tours (FR). Ennio Stipčević has provided early Croatian prints. They suit the Renaissance framework of the materials presented here.[4] The number of recordings of Croatian music found in *Europeana*'s audio materials is steadily rising, too, although most early recordings Croatian libraries (as represented by Europeana) do not contain Croatian music. The materials offering in December 2017 are in general not Croatian. A recent search found Verdi, Bizet, and Croatian renditions tangos and foxtrots. Finding specifically Croatian music composed before 1991 in North American and Western European libraries is difficult, but for a different reason: catalogues are not retrospectively converted to reflect current national nomenclatures. Croatian libraries may be able advise those outside the country as to how to facilitate retrieval of earlier materials.

An interesting reflection on the relationship between digital academic projects and revenue from commercial digital products is that they run in parallel. According to statistics gathered by the European agency Statistica, the United Kingdom leads in revenue in digital publishing, while Germany leads in revenue from digital projects related to health care. France was leading in the percentage of digital objects available in Europeana.[5] This could provide applicants for academic funding with an argument that might be understood by bureaucracies controlling grant funding. Those designing new projects can get some sense of the vast scope of existing efforts in much of Europe by perusing the European Association for digital Humanities (EADH) website.[6] Quite a few current listings focus on tools for use by others.

Topography

Digital projects have colonized wide swaths of the humanities landscape. Some that are very useful to musicologists are not exclusively focused on music. Some important ones are most readily found in aggregations, such as Germany's *Virtuelle Fachbibliothek Musikwissenschaft* (ViFaMusik).[7] ViFaMusik has so many component parts that one could spend days developing a comprehension of what all can be found there.

---

[4] See http://ricercar.cesr.univ-tours.fr/3-programmes/EMN/Croatie/index.htm. All URLs used here are permanent ones (PURLs). Material found previously should still be available today.
[5] See http://europa.eu/rapid/press-release_MEMO-09-366_en.htm?locale=en.
[6] See https://eadh.org/projects.
[7] See https://www.vifamusik.de/startseite/.

What can be daunting is the rapidity with which a global view of priorities among projects and viable means of pursuing them can change. When our research center used to published a yearbook of current applications (the *Directory of Computer Assisted Research in Musicology*, 1985-2005), research was conducted using mainframe- mini-, and early personal computers. Many projects were pursued entirely by one person without institutional support. (The humanities were not considered worthy of attention from computer scientists.) This view of three decades ago could not be more different today. Current offerings are dominated by a few large websites, most seated in national libraries. Most are in Europe, although the Library of Congress in Washington has an ever-growing array of digital treasures. The cultural strengths of respective geopolitical bases are usually evident in the presentation. Some representative sites containing large quantities of musical manuscripts and early prints, are listed in alphabetical order of their national sigla in Table 1.

<Insert Table 1>

Table 1. National libraries (selective list) with large quantities of digitized music and related items.

Although SLUB (Dresden) is a state and university library and the Biblioteca Musica (Gaspari Online) is hosted by a civic library, the first (in common the ViFaMusik, Gallica, and many other national sites) have far more non-musical resources than musical ones. The other, in contrast, is purely musical but its range of coverage mirrors the Gaspari catalogue on which it is based.[8]

What these project have in common

All of these projects are alike in containing vast quantities of material, in providing open access, and in having some easily understood system of organization. All consist primarily of scanned images. The differences involve many aspects of each undertaking. In terms of searchability, it is hard to improve on Gallica's home-page search. At this writing, its 4.2 million items include *trouvère* poetry, Machaut masses in illuminated manuscripts, operas from the wedding of Louis XIV, vast amounts of eighteenth-century printed music, and countless concert reviews from the nineteenth and twentieth centuries. Despite these riches, those pursuing French material may be surprised to discover that the University of Texas at Austin has an inexplicably large number of Lully materials, while Debussy is well served by the Morgan and New York Public

---

[8] Gaetano GASPARI, *Catalogo della biblioteca del liceo musicale di Bologna*, 5 vols., Bologna, Libreria Romagnoli dall'Acqua, 1890-1943.

Libraries in Manhattan. (While all images cited above have been digitized, not all are currently found online.)

This complex of sources near and far from their point of creation represents a widespread phenomenon: musical materials have been traveling for centuries. We can see evidence of it through the other end of telescope by looking for non-German materials in the Bavarian State Library's ViFaMusik project and in nationally supported projects in other countries. The full scope of ViFaMusik covers heterogenous materials—images, libretti, royal collections, and so forth. One of its best-known components is the notably Schrank II manuscript collection in Dresden. It contains eighteenth-century works (often in copies by Johann Pisendel) of works by Vivaldi, Albinoni, and many other luminaries of the time. As adjuncts to the music, the Schrank II website makes available a downloadable spreadsheet of its holdings (with shelfmarks and other pertinent information) plus links to watermark and handwriting databases.[9] These are highly desirable additions, partly because they demonstrate digital possibilities for source control that far exceed those of the analogue era. Despite the recency of its availability, a long string of important additions and refinements to music bibliography have been made on account of it.

A third stunning model of digital riches is available through the Swiss e-codices homepage. (E-codices identifies the Virtual Manuscript Library of Switzerland.) Switzerland's geographically intermediate status within Europe means that the origins of its holdings are diverse. Its political neutrality means that its holdings have been well preserved while analogous sources, held in countries subjected to wars of the past century, have been destroyed or, at the last, confiscated to new, unknown locales. E-codices is especially rich in medieval materials including illuminated holdings of materials culled from antiquity but conveyed through the practices of the ninth through the sixteenth centuries. Owing to the vigor of the project, e-codices now includes digital access to Swiss manuscripts in the Jagellonian Library, Krakow, the Bibliothèque Nationale of Mulhouse, and growing list of other institutions. To this the institution has now added links to dispersed parts in Cleveland, Paris, and St. Petersburg. In its entirety, the enterprise projects a strong sense of what is possible in the digital world.

One of the many benefits of using online reproductions of music manuscripts is the vividness with which they lure anyone with browser access to recapture the experience of working

---

[9] The watermark data cross-linked to SLUB holdings are found at WZIS (http://www.wasserzeichen-online.de/wzis/index.php), a pan-German all-subject database under development in Stuttgart. A lexicon of images is used to find clusters of sources on the same paper.

face-to-face with rare materials. Sometimes this means instantaneously understanding how poor the quality of preservation is (e.g., a manuscript from Claro, CH, at http://www.e-codices.unifr.ch/en/searchresult/list/one/clap/Martirologio1). At the other extreme, the delicately decorated leaf of Machaut poetry in the Bibliothèque Nationale de France found on folio 930 at http://gallica.bnf.fr/ark:/12148/btv1b84490444/f930.item reminds us of the exalted status of patrons of the actual physical objects. In Machaut's case, these figures would have been kings, bishops, and courtiers.

Cross-searching between major collections

For users in Croatia, meta-search engines such as Europeana (europeana.eu) have much to offer. Our search for Croatian materials turned up relatively little, but the collection is constantly growing. Among audio materials, we found only three items, two of them duplicates of one another. Gallica (the source of two of Europeana's Croatian recordings) provided a generous list of hits for Croatian music, the vast bulk (2,500+) being books. ViFaMusik provides listings for recent publications of Croatian music but most are under copyright. Although a search of e-codices turned up only one item representing Croatia, it is a compendious set of property documents (*Liber bonorum et iurium Castropolae*) itemizing holdings of the lords of Pula in the early fourteenth century in Istria and Venice.

Taxonomy

Topically organized website offer a very different view of available resources. The world appears very differently from the perspective of topically organized websites. The difference is worth exploring by those intending to develop new websites. The better choice will usually be the one most congenial to the materials at hand, which may originate from single users, from small groups of scholars with a common interest, and from faculties with rich troves on data on a specific topic.

What can be confusing to the user is the manner in which hierarchies of materials are structured. In Figure 1, the topic headings Books, Manuscripts, and Audio that appear in one website are then segmented into subject headings (e.g. music, literature, history). In contrast, another digital library of roughly comparable size may prioritize Music, Literature, and History above the details of holding type. This is an over-simplification, but the intent is to show how users must sometimes imagine possible organizations to be able to effectively navigate a website. This uncertainty of organization structure confers great value on a search box, such as that used at Gallica and other (but not all).

<Insert Figure 1>

Figure 1. Hypothetical model of organization of digital resources at a national library.

The landscape of digital humanities is constantly undergoing reconfiguration. Changes reflect practicalities, institutional preferences, and fluctuating access to technical expertise.

Medieval and early music on the internet

Without doubt, chant and late medieval music constitute the most popular topics in digitization projects to date. Early liturgical music is attractive to developers because of the simplicity of the music and the fact that the brevity of examples allows for long lists of holdings to be generated quickly. Eastern Europe seems to be generously endowed with illuminated missals and other liturgical books. Bibliographical control comes with difficulty because of the abundance of exemplars and the copious metadata that can be applied to differentiate practices from place to place. The Cantus network of projects links sources in numerous countries to a common index developed by Ruth Steiner, David Hiley, and many others. A different approach is evident at the University of Tennessee at Martin (http://www.utm.edu/staff/bobp/vlibrary/mdmss.shtml), where an extensive listing of textual sources containing liturgical chant has been assembled. What it offers is effectively a cross-view of digitized sources, not all of them issuing from Christian tradition but all of them medieval. Another tangent on digital resources for chant study can be found at the Cantus Fractus website (http://www.cantusfractus.org) in Trent. Music in the Feininger collection is examined for "rhythmic and proportional, actual or hidden elements in plainchant" (acronym: RAPHAEL).

Such quantities such as those under these rubrics necessarily call for finding aids. John Stinson's Medieval Music Database (original site: http://arrow.latrobe.edu.au/store/3/4/5/4/2/public/MMDB/index.htm), which is focused on fourteenth-century Florentine repertory, offers an exquisite model for liturgical repertories even when they are not tied to a particular set of sources. The database and its musical incipits can be browsed by feast, composer, or text; searched by melody; and studied by genre. Works can be arranged by temporal or sanctoral cycle. Its carefully prepared metadata is useful in a wide array of non-medieval studies. Its notable effort (in *c.* 1987) to reproduce neumes in colored notation (reflecting the original sources) was a dazzling achievement, but rapid change in hardware and software diminished the feasibility of desktop setups. The project has not been superseded, but the hardware on which it was developed is obsolete. A web display preserving the colored neumes obviates any need to reinvent this capability.

The Digital Image Archive of Medieval Music is the most comprehensive and perhaps the best known of early-music databases. Started by a triumvirate of scholars (Margaret Bent, Julia Craig McFeely, and Andrew Wathey) two decades ago, DIAMM (https://www.diamm.ac.uk/news/welcome-our-new-site/) remains a work-in-progress partly because its model is so elastic. Generous quantities of metadata are provided for every entry, but since the emphasis is on images (often available under license at high resolution), it offers a worthwhile model of coordination between data and graphical substance. The aim is to bring reproductions of all surviving sources under one system of image management. Sources come mainly from the 14th-16th centuries. Andrew Hankinson is the current manager of the site at Oxford University.

Difficult-to-categorize websites

For those seeking models to follow in new projects, there is much from which to choose that lies outside the simple coverage offered here. Composer websites rank high in most project listings. Reproductions of letters, concert programs, drafts, and other artefacts of a composer's or an institution's history are also popular. For both heterogeneity and impact, the Aaron Copland Collection at the Library of Congress (http://www.loc.gov/collection/aaron-copland/about-this-collection) ranks high. The web may be the only medium suitable to cover the composer's varied life (1900-1990). Faded school pictures show us a small boy indistinguishable from his peers. His sketches for well-known scores are self-explanatory. Recordings of early works acquaint us with the esthetic limitations of the medium decades ago. Random items, such as the typescript of a lecture on film music he gave at Columbia University in 1940, give us a glimpse of the Hollywood studio environment before synchronization of reels and recordings. In the composite, these materials offer a profile that words alone cannot convey.

Another anomalous site is that of the Beethoven Werkstatt: (http://beethovens-werkstatt.de/demo/index.html). Its aim is to reconstruct the composer's compositional process step-by-step by showing tabbed views of sketches in order of their creation. (A preview is available at the website.) The overview (Seitenübersicht) presents 18 sketches for the first movement of the Piano Sonata in C Minor, Op. 111. Encodings of scribal detail are shown under the right-hand tabs. A much simpler but idiosyncratic home page is demonstrated by the search form at César (Calendrier Électronique des Spectacles sous l'Ancien Régime): http://www.cesar.org.uk/cesar2/people/index.php). Its comprehensive categories of theatrical personnel (21 of them), its span over two centuries, and its tight coupling with precise dating

information is enormously valuable in tracing individual careers of singers and actors, profiles of troupes and theaters, and much else. Among the 80+ databases at DRM, this is one of most approachable. (Other categories are devoted to specialized search engines, image repositories, large online corpora of classical texts, and websites devoted to musical instruments. One is Polish Folk Instrument Collection [http://instrumenty.edu.pl/en/], which is both extensive and well documented in photographs and recordings.

Musical Substance

Because digitization has made such sweeping changes in the easy availability of source materials, we are inclined to overlook that possibility that our profession requires many other kinds of resources. National libraries may not be suitable places for their situation. Digitized scores are not immediately searchable. The searching of musical content belongs to a still more demanding realm of encoded music. This involves symbol-for-note substitutions, plus a broad range of detailed indications to make syntax recoverable (for example, to categorize musical genres or harmonic procedures). These arts are in their infancy.

"Born-digital" editions are often the first step in creating a corpus of searchable music. In two adjuncts to DRM—Archive of Digital Applications in Musicology (ADAM: adam.ccarh.org) and Electronic and Virtual Editions (EVE: eve.ccarh.org)—we list past and current efforts in this direction. Some notable ones are CMME (Computerized Mensural Music Editing, cmme.org), music many composers linked to texts by one poet (Torquato Tasso) in Tasso in Music (www.tassomusic.org/), and aggregations of verified encoded scores (such as CCARH's MuseData project, musedata.ccarh.org) offer a few models. The Carl Nielsen edition (http://www.kb.dk/en/nb/dcm/cnu/download.html) at the Danish National Library is both virtual and actual. It is coordinated with the online Nielsen Catalogue (http://www.kb.dk/dcm/cnw/navigation.xq). The Josquin Research Project (jrp.stanford.edu), at which scores by Josquin, his predecessors, and his contemporaries are optimized for searches within the music, has the aim of facilitating a collective appraisal of true and false attributions to the fifteenth-century composer. JRP serves other purposes as well, since its works (which include generous amount of music by other composers of the time) can be displayed and "heard" via MIDI. Its *c.* 650 pieces can inspected for textural changes and tessituras of vocal parts. (No lyrics are included in the online materials.)

Signposts for the future

Today RISM is a best-of-breed example of a digital project in musicology, and this make the prohibition of new Croatian contributions highly regrettable. Despite RISM's origins in 1952, it has only been available as a digital resource since 2011, when the music manuscript inventory (series AII) became available online.[10] The AI (inventory of music prints) series was added last year. However, in 2012 funds to enable further Croatian contributions were withdrawn.[11] In a consortium of more than 60 countries, with a collective product in formation for more than a half century, this is an inexplicable lapse. The combination of needs (languages, repertories, technologies) that underlie RISM would have challenged the most rugged of research protocols, but RISM endures and manages to improve day by day. Its value is enhanced by the establishment of an open-data policy, which enables external uses ranging from online catalog integration[12] to statistical studies of search techniques for encoded music.[13] The number of sources RISM indexes currently exceeds one million and is expected eventually to surpass two million by 2023. In the digital future, the thoroughness and clear organization that characterize RISM will continue to matter every bit as much as they did in the analog age.

Digital musicology is new enough in public consciousness to elude many of its potential beneficiaries. Its novelty prompts new research questions. Those of us who remember the rigor of research trips involving letters of introduction and permission, passports, visas, and pockets full of mini-dictionaries rejoice in the ease with which we can now view illuminated manuscripts, composer autographs, and portraits of *prime donne* from the comfort of home. Of an afternoon, we can collect all the material needed for a research paper or download several sources pertinent to a forthcoming critical edition. Yet we need to temper our enthusiasm with the acknowledgment that much of the world's music has never been edited or recorded, much less digitized. The hard work

---

[10] See http://opac.rism.info.

[11] Vjera KATALINIĆ and Lucija KONFIC, Project RISM Croatia—the Last 10 Years, *Music Documentation in Libraries, Scholarship, and Practice* (RISM 2012 Conference Report) outline the long history of RISM in Croatia and identify important opportunities that are yet to be met.

[12] Jennifer A. WARD, The use and reuse of RISM data in libraries, *Journal of Electronic Resources Librarianship*, 28:2, 129-133, DOI: 10.1080/1941126X.2016.1167551.

[13] Today's RISM music search employs the same Plaine & Easie Code that originated at New York University in 1967, together with our own *Themefinder* search engine for music (*http://themefinder.ccarh.org*), overlaid with Laurent Pugin's visualized search keyboard, and further implementation by Magda Gerritsen and others on the ViFaMusik staff at the Bavarian State Library. A preliminary study involving US RISM data (Craig Sapp, Yi-Wen Liu, and Eleanor Selfridge-Field, "Search Effectiveness Measures for Symbolic Music Queries in Very Large Databases," *ISMIR 2004*, available online at multiple sites) investigated a dozen search strategies.

of framing new questions that this access facilitates and the pleasures of forming new overviews of topics that were barely pursuable a generation ago call us forward.

Bibliography

Gaetano GASPARI, *Catalogo della biblioteca del liceo musicale di Bologna*, 5 vols., Bologna, Libreria Romagnoli dall'Acqua, 1890-1943.

Vjera KATALINIĆ and Lucija KONFIC, Project RISM Croatia—the Last 10 Years, *Music Documentation in Libraries, Scholarship, and Practice* (RISM 2012 Conference Report).

Craig SAPP, Yi-Wen LIU, and Eleanor SELFRIDGE-FIELD, Search Effectiveness Measures for Symbolic Music Queries in Very Large Databases, *ISMIR 2004 Proceedings* (available online at multiple sites).

Eleanor SELFRIDGE-FIELD, *Il Maeftro di Mufica* [sic] or Quality Control in the Virtual Library, *Fontes Artis Musicae* 62 (2015), 62-77.

Jennifer A. WARD, The use and reuse of RISM data in libraries, *Journal of Electronic Resources Librarianship*, 28 (2016), 2, 129-133.