

Notes from Du Chemin @ Haverford, February 3-4, 2012

Use these notes as a guide to the various presentations and conversations. You can listen to (or download) a sound recording of each session. See the Editor's Forum, where you can also find links to some of the resources discussed and view PDF files of the individual presentations. The conference packet is mentioned in many of the discussions. This, too, is available via the Editor's Forum.

If you would like to respond to any of the materials assembled here, please use the appropriate page of the Editor's Forum (either the Du Chemin @ Haverford page, or one of the thematic pages).

Session I. Getting into the Music.

- Freedman: an overview of the Lost Voices project, its methods and aims.
 - Since 2009: Image archive for partbooks, PDFs of modern editions, commentaries, instructions, database and interface.
 - Now 2011-2013, with ACLS, NEH, and CESR funds: dynamic editions (*show* the variants), reconstructions (for discussion and debate, and study of style), materials for the literary texts (for understanding of chansons, and the broader cultural currents in which they participate, and models for collaboration (via standards for sharing files, methods; and also in our reconstructions: 80 incomplete chansons, rebuilt with an ear for style and eye for engagement with each other and with theoretical sources of the 16th century).
 - Conventions will be an important part of our reconstructions, since they offer a constrained vocabulary of compositional choices. These will form the core of our thesaurus for this repertory.
 - The reconstructions are meant as thinking pieces, not as an attempt to recover original artworks. We want to expose what we know about style, and how we know it.
 - We listened to some examples (Claudin de Sermisy and Clément Janequin), then turned our attention to the Thesaurus and how might be organized. In brief: through empirical observation and through dialogue with practical/theoretical writings of the period. We began with the three broad headings set out in Cerone's El melopeo (1618): cadences, accompaniments, and imitative duos. But of course the details are more interesting than the general categories. We looked mostly at examples of the first two, then paused for discussion.

Discussion

- Julie Cumming: All these situations can be understood as 'contrapuntal'. This will be important as we advance to implement various electronic tools for the study of polyphonic music. A related point will be for us to distinguish between 'presentation types' (by which Peter Schubert and his team mean the typical patterns of repetition or texture) and 'interval patterns' (which are shared across the presentation types). Julie also observed that some of the 'special' sub types of cadences in Freedman's inventory in fact have a long history in the 15th and 16th centuries. Exs: 'non-double leading tone', of the sort found in the Menehou example cited by Richard. Also: 'cadence hidden within a cadence', as when we have one cadence that yields into a second that is a fifth below. We need to mention the theoretical sources that present these seemingly unusual combinations and explain their history.
- Cynthia Cyrus. Rhythm matters. It is essential to formation of cadences. We need to describe some of these conventions, too, as a way of showing not only cadences that do take place, but ones that might have taken place. Cynthia also later suggested constraining the Thesaurus to no more than 30-35 basic possibilities in all.
- Susan Weiss: we need basic data about the repertory, such as ranges, mode. It was also suggested that we need statistics on rests, that is: how often are some voices silent, which will tell us about the prevalence of duos, trios, etc.
- Patrick Macey suggested that search tools would also be able to tell us the probability of things like leaps of a certain interval.
- Jesse Rodin: We will also need to move beyond the lower level descriptions to the patterns implied by these statistics. What are the markers of a particular gesture? Where does it occur in a piece? What is its function? The idea of 'collocation' seems important here.
- Stefano Mengozzi. If compositional choice is the aim of the Thesaurus, then we need some more subtle than 'must happen' or 'must not happen' in a given context. Instead, we need to model what Lerdahl called

'preference rules', but in this case for 15th and 16th century music. It's obvious that the language for the Du Chemin chansons is quite closely constrained. For fifteenth-century repertoires we need something more agile.

Session II. Perspectives from the Library.

- Freedman: The Du Chemin albums and the Thesaurus imagined as a collection of digital objects. They contain musical information, but no less important will be the information about this information that will matter. We need to be able to identify authors, contributors, sources, versions. And we need to be able to cite and juxtapose passages are part of our reconstructions, editions, and arguments. For this we need Librarians and a robust Content Management System.
- Adam Crandell. Two key challenges: Discoverability and Organization, both Internal and External.
 - External: We know that the editions, variants, thesaurus entries, etc, will have metadata associated with them (name of editor, contributor, author, source, and the various elements of the thesaurus, too). But which of these metadata do we want to be harvestable but those beyond our site? How 'deep' a data set should be available to outside search engines? To the smallest level of musical detail? Or should we make the 'authors' of our reconstructions something that can be found through an internet search? Should someone be able to learn that (for instance) Peter Schubert's work has been cited in our resource? And what happens as our own roles change (from editor to commenter to analyst to reconstructor)? Our collection is a dynamic one! What should be available via resources like the Open Archive Initiative?
 - Internal: Musical information and the data about it are now all to be in XML. So they are equally available to us as search terms. For this, however, we will need to have a controlled vocabulary (that is: the Thesaurus, in addition to the larger scale terms of composer, genre, reconstructor, analyst. And yet we also need a space for free form comments as part of the open discussion of solutions and challenges. These comments will have their own metadata (contributors and subject matter, since they are attached to certain pieces or concepts).
- Joe Gilbert. As part of development of any web-based project for a broad set of users, we need to build a Wire Frame. It helps us to focus on the conceptual issues of the project (and not the color or font, which come later). Here instead the focus is on layout, flow, and the other key functions we need. Joe explained the current concept of 'companion' sites: The Lost Voices Project (with thesaurus, reconstructions, etc), and the Chansons Nouvelles Resource (essentially an archive of images, editions, commentaries).
 - The aim is to make them distinct (so readers will know which one they are using) but deeply linked, so users can jump back and forth as needed. The first views, however, need to draw users in so that we balance the needs of *easy* and *deep* engagement. Important: *acknowledge and expose method right at the outset*.
 - Next steps: we need to define the '*content modules*' for the projects, listing all attributes and elements: editions, sound files, images, composers, phrases, thesaurus categories. Some of these can grow as the repertory grows, but we need to know the types of elements we are defining at the outset, so we can relate them to each other. Then we need to specify the missing technologies.

Discussion:

- Michelle Oswell: Can we search by 'contributor'? This should be part of MEI. It could make the resource useful for those who want to claim credit for their work.
- Cynthia Cyrus: We need to be able to recover a particular view or object, and cite it for others to recover. We need a 'cite this work' and 'cite this entry or comment' link. [See p. 48 handout: that is the concept of saving to favorites. If it can be saved here, it can be cited and exported as a link.]
- Joe Gilbert: Here we probably need to revisit the notion of 'work'. Does it include "all versions" ? By all authors?
- Darwin Scott: The Thesaurus is essential to this. Without a controlled vocabulary for Renaissance music, you will not be able to recover the local moments that inform the style, and inform the reconstructions.
- Catherine Motuz: We also need some way to search for, or at least mark up, the positional aspect of these thesaurus elements. Where do they come in a piece? The principle of 'collocation' seems important here. This could be something done as part of analytic mark up. Or could be built in to a search engine: find X "near" Y. Also "not near". Or "where in relation to what"? Statistical information about this could help.

Session III. Music Encoding.

- Freedman started with an overview of current situation. In brief: how can we move from Sibelius, Finale, and other practical notation programs to systems of encoding that will be open to all sorts of computer-aided analysis; and how can we do so in ways that will keep the all-important sets of meta-data closely associated with our work (so we can identify composers, editors, reconstructors, sources, variants)? The conference packet shows some of the basic steps we can take right now towards these goals by inserting essential information in various fields already found in proprietary software, then exporting these as MusicXML as an interchange format. But there must be a better way . . .
- Andrew Hankinson explained the history and profile of the problem. Sibelius, Finale, and other programs are good for _graphical_ presentation of music. That's useful for musicians who want to read scores, and the editors who want to create them. But computer-aided analysis of music (whether we're talking about variant readings or sophisticated search and analysis) will depend on *structured data*.
 - Andrew offered a survey of different systems that have emerged over the past two decades for preserving musical information in structured sets. These are likely to remain quite separate from the need for representing musical information for players and editors. But there is some promise of graphical end-user interfaces that will allow editors to add/update/correct rich data sets in ways that will combine graphical and structured approaches. And with XML-style systems, we can maintain a complete record of editorial interventions ("track changes"). Space will be no problem, and we can have as many versions of a text as we like. In order to support these systems
 - Support systems will be crucial for the next steps: content management, search and browse features. Presentation tools that will allow us to edit on line and see/compare multiple versions of a 'work'
 - Formats: MusicXML will work as an interchange format, but nothing more than this. MEI has much promise as a 'relational' XML system, but as yet there is no software.
 - Challenges: how to manage multiple files (version control)? Work flow (how will users submit their contributions? Can we automate the process? What is a musical query, anyway? We need to move beyond relatively simple searches for strings of pitches or rhythms. There are more sophisticated options like HumDrum and Music21, but these are all for the pros, and all require precomputation of answers to our questions.
 - How will we develop tools for different notations (mensural, modern, avant garde)? How will we manage 'crowd sourcing' for tags? What means to verify the 'authority' of a contributor to tag our files.
 - Examples: **Papyri.info** This can serve as an example of how various users contribute to a collective 'puzzle'. It allows them to transcribe, annotate, and make comments about fragmentary text while preserving the meta-data about who wrote what when. It takes a standard system of annotation and editorial mark-up "Leiden" and transforms it as TEI.
 - Montreal projects: Digitized and Encoded liber usualis, using MEI and OMR. Now fully searchable by neume no less than text or pitch. See <http://ddmal.music.mcgill.ca/liber> Also viewer **DIVA**, allows viewing of different images and digital objects. Allows 'cite/save/export' of individual images. This would work well for Du Chemin or other facsimile projects. Open Source. See sample project: <http://ddmal.music.mcgill.ca/salzinnes>

Session IV. RoundTable of Digital Projects.

- Freedman began with a brief response to a recent series of articles by Stanley Fish, in which various anxieties and doubts Digital Humanities were cast in broad terms, spelling what seems (to Fish) to be the end of scholarship and in turn the end of scholarly authority. Freedman: yes, new technologies have often worked in disruptive ways. Even print was regarded with both hope and dread during its first century. We need to be alert to the concerns of colleagues, and explain what we are up to, and why it matters.
- Jesse Rodin reported on the Josquin Research Project at Stanford. They are encoding music of the fifteenth and sixteenth century. They want to analyze large numbers of pieces to learn about norms, and also to address questions of style and attribution. They are interested in using computers to discover the grammatical structures of the music.

- At present they begin with established editions of music (they are not doing critical editing), which they enter with Finale. This in turn is marked up to reflect original mensuration, and then transformed via MusicXML to HumDrum.
 - This in turn allows them to pose questions dynamically: they can search for melodic and rhythmic strings, produce PDF's of any section of a piece for display, with highlights of query results. They can create different kinds of visualizations of the patterns they discover. This can be done by piece, and by corpus.
 - Grad Sourcing is key to their approach but they are also accepting submissions of previously encoded music.
 - As yet, not concerned with texts, or with producing editions in an on-line environment.
- Xavier Bisaro. An overview of the Atelier virtuelle de restitution polyphonique, CESR, Tours.
 - It begins as a database of music from the 15th, 16th, and 17th centuries with missing voices or sections. And it will grow to articulate the methods by which we will reconstruct these missing materials: contributors will cite the reasons for their decisions, and these will be part of a dialogue about the material. The Thesaurus is key to this, since it is the basic resource for style and statistics about particular composers, genres, and periods.
 - The restorations are not meant to be definitive, but rather collaborative. Recognizing that these will be "open" works, they are interested to find ways to present versions, add comments.
 - The Ecole thematique (2010, 2011, 2012) is an important public manifestation of this work, since it allows time for students, scholars, and performers to gather for open discussion of the music and the methods.

Discussion:

- What about the place of old notational systems in all of this? Do we blind ourselves to certain kinds of relationships when we encode modern transcriptions rather than mensural (or even modal) systems directly? Will it be possible to link back to original notation in some way?
 - Encoding for notation and for analysis are not necessarily the same thing.
 - The limited capacity for meta-data in programs like Sibelius and Finale presents big problems for work like this, but standards like MEI provide ample room for customized mark-ups and also links to facsimiles.
 - For the Thesaurus: how will we define the elements and select them? Are we only looking for conventions, or could we also tag things that are unique? Will it be possible to do some of this automatically at some point? That relationship to the theoretical literatures of the period? Will elements be confined to sections of 'real' compositions? Or could we also imagine hypothetical structures proposed as archetypes by modern scholars?
- Julie Cumming and Ichiro Fujinaga. SIMSSA, and ELVIS. Searching, Analysis.
 - SIMSSA: a "Single Interface for Music Score Searching." Towards GoogleBooks for music: a unified search engine that will allow searching through an entire universe of scores, made possible through OMR (optical character recognition) and MEI (which will understand the notation in more than symbolic terms. It will provide tools for a DIY approach to the review and mark up of results, since it will also be compatible with various other tools, like MuseScore, DIVA viewer, HumDrum, and Music21. Sources will draw from the IMSLP open source library, DIAMM, Early Music online. They will also be looking at musical notation that appears within non-musical books, an oft overlooked source of musical works and ideas that would be very important for scholars. See PDF of presentation for more information.
 - ELVIS: An "Electronic Locator of Vertical Interval Successions." They are interested to build a sophisticated engine for searching musical repertoires--something beyond the usual searches for strings of pitches or rhythms. They will do both corpus studies (seeking out patterns across thousands of pieces) and also detailed studies of individual styles and pieces.
 - Their latest thinking is much informed by ideas of 'collocation', which is to say the insight that languages are often heavily constrained depending on context, and that one or two choices often imply others.
 - For counterpoint, they will look for pairs of intervals, building on the basic vocabulary of successions of intervals suggested by theorists like Tinctoris. From these they

will look for repeating sequences. The pairs function as words, and from these they will be able to 'look' for cadences, stretto fugas, etc. See the presentation [PDF](#) for examples of how they are imagining their work. They will begin with scholars and students, then move to machine assisted work.

Session V. Text Encoding.

- Freedman: an overview of the opportunities and challenges presented by the literary texts of the Du Chemin chansons. To date, the CESR team has made a database of all the chanson texts, transcribed line by line, with basic information about rhyme and versification. They have regularized spellings, much as we have done in our modern editions of the chansons in Sibelius.
- But now we want to move towards the next steps: integrating these texts with the Du Chemin archive so that readers can study the complete text while they consider the music. We want translations, and information about versification and rhyme, so these can be studied in conjunction with the music, too
- What is more, we are thinking of TEI as a way to allow us to compare textual readings among the partbooks, something that can tell us about working methods of the press and also help scholars explore problems of orthography and punctuation. See what Stanley Boorman has done for the case of Petrucci partbooks, comparing compositor styles among partbooks.
- Lastly, the TEI will allow us to consider the literary texts of the chansonniers in relation to the many non-musical sources in which they appeared in the sixteenth century. This, too, will allow us to compare versions for differences of spelling, punctuation, and also look at patterns of circulation. The Du Chemin prints and the literary prints, too, are anthologies, and as such present some quite interesting opportunities for scholars to see which texts were printed in what sequence. This matters for the Du Chemin chansons in particular, with their many chanson-response combinations.
- Freedman showed an example of what might be learned from all of this: Le Rat's setting of a text by Clément Marot, with its fascinating musical 'parenthesis' that in fact reflect punctuation in the original literary source (but are not found in Du Chemin's typography). We compared TEI editions prepared by the CESR with a digital facsimile of the original source, and we looked at the various elements we might encode from the Du Chemin chansonniers themselves. See the Conference Packet handout for he details. We sang through the piece.

Discussion:

- From online contributors Van Orden, Zecher, Rouget, Koutsobina:
 - We need information about versification, meter.
 - Translations, especially translations from the sixteenth century when available
 - We need information about chanson-response pairs, and which appear near each other, and which are mentioned in the sources, both literary and musical.
- What to encode? Everything on the page? Or do we add a declaration at the outset that we have selected only certain features (for instance, that we have resolved i/j and u/v without remark)
 - Formework? This could matter for attribution, since sometimes the attributions in the title page and in the body of the book are not the same.
 - All of the internal text repetitions, and even the "ij" markings? [Some said this was properly part of the `_musical_` work and not the literary text per se. If we encode every repetition, how would we also encode a 'line by line' version? Without this, it will be almost impossible to compare the partbooks with each other, and with literary sources?
 - Resolve i/j and u/v? [Currently the CESR TEI guidelines do this without comment. Some nevertheless thought we should preserve every original distinction in the text, then encode our 'resolved' version as an editorial intervention. The TEI display would parse this according to need for display, but would preserve the complete original for some future use.]
 - Abbreviations and abbrevigraphs? [Again, the CESR TEI manual already deals with these: they expand abbrevigraphs, since there is no ambiguity about what they represent. When there is some doubt, they encode the original and show the editorial intervention?
 - E barré? This is important, since it tells us about elisions and text placement--things not present in the literary sources.
 - Of course we also want to encode literary sources. We will need to make some reasonable selection. The more detail we encode for the musical 'texts', the fewer pieces we will have time

for, and the fewer literary sources we can control. One strategy for Phase I (summer 2012) would be to settle on two or three chansonniers and one or two literary sources--with the understanding that only a small number of 'hits' will join the two. Note that a surprising number of literary sources are already available in digital facsimile.

- We will need to move promptly to:
 - 1. Identify sources we want to encode.
 - 2. Make decisions about what we want to encode, and how. We can use the CESR TEI manual as a starting place.

Next steps

(Ideas for Discussions among Freedman Bisaro, Weiss, Fujinaga, Cumming, Thomas, Cyrus, Hankinson, Motuz, Boynton):

- Refine the Wire Frame.
 - Detailed Content Model (that is, all the different types of files for each work)
 - Cite/save/export;
 - Searches for 'collocation' models of near/with/not.
 - Controlled choices for those doing markup.
 - User roles (analyst, commenter, reconstructor).
 - Also explanation of methods and aims.
 - Anticipation of ways to integrate the Example Viewer (see below) with the remainder of the Lost Voices.
 - "Promote Work" feature, that will draw readers in to a focused discussion around a single work or issue.
 - Harmonize browse and work view for the archive.
- Build Example Viewer
 - Explore VexFlow from Server. What programming required? What sorts of information would it need to render any given 'span' of measures from MEI to HTML 5 frame or window, surrounded by selected information about the piece, source, and category at hand?
 - References: 'measure to measure', or a 'push pin' model? That is: will viewer render a precise span (surrounded by some variable number of 'context' measures? Or will we simply indicate the 'start' of a particular element?
 - Later: some way to superimpose variants (different sources of the same work) or versions (different reconstructions of the same work). Show differences via highlights? Or some kind of expandable view?
- Thesaurus.
 - Controlled vocabulary (avoid more than 30-35 choices, otherwise users become confused).
 - "Free" spaces for comments and responses
 - Reference system: Word ID, Measure ID, XML ID?
 - Compile Thesaurus as a WIKI, which then leads back to the Du Chemin and AVRP resources.
 - Human indexing: web form to a database. Make web form (with Joe Gilbert?) for Thesaurus entry--start with poetic text (thus phrases), then attributes/categories observed in each. This would not preclude reuse of the small-scale tags in other arrays or contexts.
 - Double or triple "key" each analysis: to confirm/refine the methodology, have then train together, but work independently.
 - Machine indexing: can come later as the tools emerge and as the human indexing refines what we want to search for.
 - WHO defines the categories? We need to encourage experts on music theory to help refine the categories. Then train graduate assistants as analysts, and then as reconstructors.
 - Citations to secondary literature.
 - Thesaurus Sample by July 2012: three books of Du Chemin chansons nouvelles analyzed. Some sample of practical/theoretical sources (like Lusitano, Cerone, Sancta Maria).

- Reconstructions by July 2012: 1 dozen pieces? Encourage 'competitive' reconstructions, with requirement that they cite elements in the Thesaurus. Viewer to show/compare the results.
- All in time to demonstrate Fall 2012.
- Prepare Files for Sharing.
 - Our groups should work on some basic principles for sharing data.
 - Meta data in Finale-Sibelius (and how they persist in MusicXML and MEI)
 - Naming conventions for files
 - For Du Chemin: we need to delete incipits, fix mensuration signs and repeats.
- Texts
 - Freedman and Bisaro to confer with Carla Zecher and BVH team to discuss editorial policies and encoding standards.
 - For texts to be encoded: pick the same books as we will use in the Thesaurus? Livres 1-3?
 - Pick literary sources (by what criteria? things on line, or with maximum overlap with each other? with Du Chemin?)
 - Three stages: 1) basic transcription, using superius. 2) detailed mark up of differences, repetitions, punctuation. 3) define schema and XLST. Possibly: find some kind of Leiden style editor as use for Papyri project?
- Programs/Events
 - Boot Camp for refinement of Thesaurus categories: two or three of us convene (around RSA meeting) to agree on some of the problem areas and work them out. Also to build citation bank of primary and secondary sources that illustrate/explain the categories.
 - Boot Camp for analysis: to train students, and identify those ready to work on reconstructions.
 - Presentation at Med Ren 2012? Freedman has been asked to offer some kind of keynote, but there is still time to submit proposals for presentations. They could all be mentioned in the keynote.
 - AMS (or other) 'pre/post conference sessions'. That is: use some of our grant funds to extend the stay of participants for a working afternoon or evening. This could be used to draw in new participants, or otherwise take advantage of an assemblage of colleague and students.
 - Newberry Library 2012-13. Plan an event something like Du Chemin @ Haverford, but invite colleagues with help from the Renaissance Studies Center at the Newberry.
 - CESR Ecole thematique III, October 2012. Invite European colleagues. Also consider bringing north american colleague (a scholar and one or two students).
 - Teams of visitors or guests to do a kind of 'road show', visiting each other's institutions.
 - Pedagogical materials: sample assignments for transcription, analysis, recomposition. "The Pedagogy of the Open Work" NEH funds cannot be used for textbooks per se, but our project does anticipate the need for involvement of advanced students at different stages.
 - Recordings/Performances. Again, not with NEH funds, but these seem a natural component of the practical work at hand.