LE STUDIUM
Loire Valley
Institute for Advanced Studies

LE STUDIUM Multidisciplinary Journal
ISSN 2669 – 7807

www.lestudium-ias.com

FELLOWSHIP FINAL REPORT

Making Musicology in a Digital Age

Richard Freedman[1], Philippe Vendrix[2]

[1]Haverford College (USA),
[2]CESR, Université de Tours, France

REPORT INFO

Fellow:
Richard Freedman
From Haverford College (USA)
Host laboratory in region Centre-Val de Loire: CESR, Université de Tours
Host scientist:
Philippe Vendrix
Period of residence in region Centre-Val de Loire: January-May 2025

Keywords:
Musicology, Renaissance Studies, Data Science, Digital Humanities

ABSTRACT

What does the study of music look like in an age of ubiquitous data? How have we collected and categorized music, in all its richness? And how can digital tools help us understand musical style and structure, particularly when it comes to the music of the distant past? As my host Philippe Vendrix explains in the introduction to EarlyMuse (https://earlymuse.eu/; a major European Cooperation in Science and Technology [COST] project), the field of early music has always been marked by hybridity: a meeting ground of theory and practice, of scholarship and performance, of the material and the immaterial. Now the digital turn in scholarship, and the central place of digital media in culture more broadly, bring new opportunities and challenges to this mix.

The experimental basis of my project, briefly stated, is to find ways to advance two key axes—one technical, the other social. Thus, one portion of my work will focus on "music as data" and "data about music" at both micro and macro levels of detail. During this phase of work, I will advance techniques for the curation of quality data, the development of algorithms and machine learning systems for music, and the elaboration of systems for mapping and linking knowledge. But I will also help seasoned professionals and aspiring graduate students alike direct this kind of "applied human intelligence" to the machine tools that increasingly mediate our experience with all forms of culture, and that shape knowledge systems, too. If musicology is to succeed in a digital age, we will need to involve digital natives in this work. And if young musicologists are to succeed in the rapidly changing landscape of the academy and culture industry more broadly, they will need to hone digital skills early in their careers.

1- Introduction

New technologies of writing have always complicated the relationship between authors and readers. This is especially true of the performing art of music. From the beginnings of western musical notation, to the advent of the printing press, to sound recording, and now to the digital domain, new technologies of transcription brought about means for controlling the effects and purposes of music, even inaugurating a new sense of it as intellectual property. Each was a "new medium" of its day, and each brought with it new ways for composers, performers, and listeners to interact around musical ideas. Now it is transforming scholarship, too, particularly with the arrival of Large Language Models and the

uncanny knowledge they seem to reveal.

My time in Tours during the Spring of 2025 advanced these concerns via a number of different initiatives underway the RicercarLab (https://ricercarlab.cesr.univ-tours.fr/), the musicological research and publication division of the CESR, Europe's leading institution for Renaissance studies. Some work was highly detailed and technical, involving the elaboration workflows for musical data and metadata. Other efforts were aimed at the development of communities of practice, both among the students and researchers of the CESR, and internationally, through the COST-Action EarlyMuse project directed by Vendrix.

2- Experimental details

One key center of concerned technologies of inscription. From the beginnings of western musical notation, to the advent of the printing press, to sound recording, and now to the digital domain, new technologies of transcription brought about means for controlling the effects and purposes of music, even inaugurating a new sense of it as intellectual property. Each was a "new medium" of its day, and each brought with it new ways for composers, performers, and listeners to interact around musical ideas. During the last 10 years (and particularly in a series of collaborations that I have undertaken with the CESR through projects like *The Lost Voices Project* (https://digitalduchemin.org/), *Citations: The Renaissance Imitation Mass* (https://crimproject.org/), and *Gesualdo Online* (https://ricercar.gesualdo-online.cesr.univ-tours.fr/) we have pioneered new kinds of dynamic, digital editions made possible by The Music Encoding Initiative (https://music-encoding.org/), a XML-based standard for the scholarly representation of musical scores of all kinds.

MEI and its related technologies are now widely used by scholars. But there are many significant roadblocks to the *efficient* use of them, particularly since the various music editing platforms and applications in common use (such as Sibelius, MuseScore, and MEI Friend) leave various 'last mile' gaps in the final results, particularly since the rich features for annotation, variant readings and other scholarly markup required in editions of early music are tedious or impossible to craft with standard software. And so part of my time was spent developing a set of freely-available Python tools ('MEI Tools'; see https://github.com/RichardFreedman/mei_tools) that smooth the editorial process from desktop to screen. Editors can now record their annotations directly in their graphical programs, then follow up with robust post-processing steps in a single library. Hundreds of files can be updated at once and readied for publication in digital form.

Another part MEI Tools also handles the other side of the 'last mile' equation: in this case the curation of the essential metadata (composer, title, editor names, sources, and rights statements). Of course it is routine for databases of music scores offer users various ways to search for pieces—by composer, title, date, or source. And one of the genuine advantages of MEI XML files is that they offer not only support for scholarly annotation, but also rich metadata as part of the "head" section of the file. Unfortunately there is still no simple way to make sure that the metadata in an individual file is synchronized with that found in the databases that preserve those scores. The metadata module from our MEI Tools solves this problem in a convenient way: metadata can be harvested from an API, or even a simple spreadsheet and automatically populated into dozens of MEI files at once. Or conversely metadata from a corpus of encoded scores can be queried and retrieved as part of data archives. MEI Tools are now a regular part of the workflow for CESR/RicercarLab editors, and a key part of their plan for sustainable workflows and projects.

Another part of my time in Tours was spent in the elaboration of a Python library for the machine-based *analysis* of music scores: CRIM Intervals (https://github.com/HCDigitalScholarship/intervals/blob/main/README.md). Developed by participants in the CRIM Project (funded by major grants during the last seven years, including my previous Le Studium Fellowship),

this library makes rapid work of a wide array of music analysis tasks, putting the details of the subtle craft of Renaissance counterpoint into a big data framework. We can search for patterns in hundreds of compositions at once, returning new insights about style and structure.

During my time in Tours I refined some aspects of these tools with colleagues David Fiala and Philippe Canguilhem. In particular we showed how to derive new insights about a large set of 15th-century chansons contained in a handful of manuscript sources (the so-called Loire Valley chansonniers) for which the Ricercar team had recently obtained complete MEI transcriptions. I presented the results of this work during the Le Studium colloquium organized by visiting scholar Carlo Bosi, and offered a related workshop-presentation at KU Leuven, too (in this case focusing on a recently-discovered source from the complex). CRIM Intervals can be challenging for users new to Python and Pandas environments. And so as part of my time in Tours, I developed a web-application version (CRIM Intervals Streamlit) that makes it simple to interact with the system and return novel reports of musical events, visualizations and other information.

All of this work was in part motivated by the needs of expert scholars. But it was also brought directly into the musical classroom. I visited David Fiala's in digital humanities for master's level students of the CESR. And for ten weeks I was a guest instructor in Philippe Canguilhem's "Introduction to Digital Musicology" course at the Département de Musique at the nearby Université de Tours. It was a pleasure for me to get to know some two dozen French students. Together we explored two key dimensions of the digital turn in musicology: data about music (chiefly through an exploration of Spotify and genre metadata about the Beatles) and music as data (in this case via the tradition of the contrapuntal duo in Renaissance style and beyond, using CRIM Intervals and MEI methods). I think the students learned a good deal; I certainly benefitted from the experience, above all by teaching the material entirely in French. And now my colleague Philippe Canguilhem (in collaboration with Ricercar team members Augustin Braud and Ailin Arjmand) is repeating the experience in the Spring of 2026.

We also continued our efforts to expand the ring of collaborators associated with RicercarLab and the CESR. In February we welcomed Marcel Klinke (Assistant Professor at Heidelberg University) who is helping to direct Christiane Wiesenfeldt's project on the Imitation Mass (they are contributing editions to the joint Haverford-CESR CRIM Project). Later I was invited to Heidelberg University, where I taught a 30-hour module for a dozen graduate and undergraduate students ("Blockseminar zur digitalen Musikanalyse").

Still other work took me away from Tours from time to time, via a number of initiatives organized by Philippe Vendrix. In February I was invited to present a paper "MEI: What it is and Why it Matters", to an international gathering of editors and publishers held at the Centre de musique baroque at Versailles. I also took part in two planning workshops (one in Ljubljana, Slovenia, the other in Bratislava, Slovakia) for EarlyMuse (https://earlymuse.eu/) , a major COST-Action project supported by the European Union. EarlyMuse is devoted to the development of new research initiatives in the study of music from before 1600, and involves academic institutions, libraries, archives, and music industry groups from some two dozen countries. I contributed extensively to the working group reports on Publications and Policies, and continue to advise Philippe extensively during the preparation of the final project report for the EU commission.

3- Results and discussion

As I noted at this outset of this report, each new technology for the inscription of music has involved a transformation in the roles we enact through this art. The lessons of my time with CESR colleagues and students this year have borne this out. We *all* have things to learn from each other, and so it is both natural and productive that digital humanities classrooms and research laboratories should intersect with each other in productive ways.

No less than the development of techniques and tools, in short, the development of people, institutions, and communities of practice will be a key aspect of the digital scholarship, and I am glad to have been able to advance these aims through my association with Le Studium and the CESR. These, no less than the assemblage of scores and metadata, are keys to future success.

4- Conclusion

It is certainly too soon to draw final conclusions from our work. But thanks to the experience of the Spring term at Le Studium, I think we are well positioned to take advantage of the vast array of « data about music » and « music as data » that the CESR has curated during recent decades. And indeed, the next phase of our work (as suggested in the next section of this report) promises to make those resources newly accessible in ways we are only beginning to imagine. One key to this work will be experimentation with Large Language Models.

5- Perspectives of future collaborations with the host laboratory

The rapid advancement of LLMs in the last two years builds on decades of research into machine learning and monumental leaps in computation. Many of us are now familiar with (and worried about) the generative possibilities of these technologies, as they are used to write emails, compose or summarize prose, and even create images or videos, sometimes displaying an alarming tendency to hallucinate uncanny falsehoods. Tools used the wrong way can be dangerous or destructive, but now more than ever we must cast a critical, human eye over their mechanical limitations. Avoiding LLMs is already close to impossible. The challenge will be learning how to use them effectively to make the most of the unprecedented opportunities they offer for new modes of research. In this two-year project, we explore two areas that are focal points in our work as researchers and teachers: data about music (the structured representation of claims concerning works, performances, and the people who make them), and music as data (the structured representation of symbolic music scores).

Crossing both tangible forms (people, places, events, instruments, written scores) and abstract concepts (emotions, patterns, and even the notion of the musical work itself) the domain of music provides an ideal testing ground for the power of contextually based LLMs to help us bridge the gap between musical artefacts and how we know them (or how composers, performers, and listeners of the past knew them). Thus one important dimension of our work will be the construction of domain-specific representations of writings about music and musical scores, which in turn will allow us to traverse these two ways of knowing music in new and informative ways.

Large Language Models, for instance, offer the opportunity for creating natural language interfaces with structured data. Users can ask questions in plain, everyday language. The system then converts these questions into structured queries that work with digital 'endpoints' (API's that use SPARQL queries, for example), making it easier to retrieve information from the database. Anthropic's recently announced Model Context Protocol, for instance, details the transformative possibilities of this kind of interface between human queries and structured data, returning narrative responses from highly structured datasets.

We also note that the opposite trajectory is also possible, returning produce structured responses about some body of prose writings. That is, our tools could attempt to identify and categorize the people, places, ideas, and sounds described in these texts, and in turn present them in a format already suited for inclusion in a structured database or in tabular format for subsequent data analysis. The usually tedious creation of TEI or MEI is one area where LLMs have enormous potential to help identify and mark textual elements of interest. We could use these to vastly speed up the process of tagging documents. There are large bodies of writings suited for this work, including not only public-domain music criticism, but also historical theoretical writings about music, particularly from the Early Modern period, which will in turn relate nicely to the corpora of encoded

scores that the CESR has already assembled. Some of these writings are in English. But the exploration of other theoretical writings on music is not, and so once again we will be interested to discover what we can learn with (and about) the capacity of LLMs to map knowledge in languages other than English.

During the Spring term I started conversations about all of this work with RicercarLab colleagues, and also with both local experts (CESR professor Carlos Gonzalez Gallardo specializes in LLM research with historical documents), and other collaborators around the world (including ontologist and Le Studium Fellow Emilio San Filippo, and data scientist Daniel Russo-Batterham). This spring we will submit a proposal to the Schmidt Sciences 2026 Humanities and Artificial Intelligence Virtual Institute (HAVI) initiative for a multi-year grant to support this work (see: https://www.schmidtsciences.org/opportunity/2026-humanities-and-artificial-intelligence-virtual-institute-havi-rfp/).

6- Articles published in the framework of the fellowship

My work was made available to wider audiences in a number of venues. These include:

Public Presentations

"Close and Distant Reading of the Leuven Chansonnier", The Digital Edition of Medieval to Early Modern 'Song': Methodological and Interpretative Perspectives, 13-15 May 2025. CESR, Tours, France.

"Making Musicology in a Digital Age," Forschungskolloquium des Musikwissenschaftlichen Seminar, May 6, 2025. Heidelberg University.

"Close and Distant Reading of the Leuven Chansonnier", Musicology Seminar, April 30, 2025. KU Leuven, Belgium.

"MEI: What it is and Why it Matters", Partition et numérique: Journées d'accélération, Centre de musique baroque de Versailles, 17-18 February, 2025. Versailles, France. https://pepr-iccare.fr/events/partition-et-numerique/

Peer-Review Published Articles

Richard Freedman, Emilio Sanfilipp and Alessandro Mosca, "Ontological Modeling of Music and Musicological Claims. A Case Study in Early Music." *International Journal on Digital Libraries* (in press)

Richard Freedman, "Music and Print Culture in the Renaissance", *Routledge Resources Online—The Renaissance World*, ed. Giovanni Zanovello and Erika Honisch (in press)

Code and Related Tutorials

Richard Freedman, "MEI Tools: Python Code for curation of Data and Metadata in Encoded Scores" https://github.com/RichardFreedman/mei_tools (2025)

7- Acknowledgements

I am very grateful to the many colleagues who welcomed me to Tours: Vincent Besson, Augustin Braud, Hyacinthe Belliot, Camilla Cavicchi, and Ailin Arjmand were my daily companions in the RicercarLab office. David Fiala and Philippe Canguilhem invited me into their classrooms, while Philippe Vendrix was an endlessly enthusiastic supporter of our work and truly strategic thinker. Finally, I cannot imagine any of this being possible without the professionalism and warmth of the staff of Le Studium: Maurine Villiers, Sophie Gabillet, Aurelien Montagu, and Marie-Frédérique Pellerin.

8- References

Published papers, public presentations and code repositories are noted throughout the body of the report.