Music Scholarship Online (MuSO): A Research Environment for a More Democratic Digital Musicology

Timothy C. Duguid <tim_dot_duguid_at_glasgow_dot_ac_dot_uk>, University of Glasgow
Maristella Feustle <Maristella_dot_Feustle_at_unt_dot_edu>, University of North Texas
Francesca Giannetti <francesca_dot_giannetti_at_rutgers_dot_edu>, Rutgers University
Elizabeth Grumbach <egrumbac_at_asu_dot_edu>, Arizona State University

Abstract
This paper describes the work to date on Music Scholarship Online (MuSO), an online research environment for digitized and born-digital music resources that inscribes itself within the federated model of the Advanced Research Consortium (ARC). With the project now in its third year, MuSO has reached an inflection point where it has developed a music-centered RDF schema and demonstrated the potential for federated searching across ARC nodes by crosswalking eighteenth-century music content from Europeana into ARC. The case study presented here outlines the dissemination role that MuSO proposes to play within the music research community, the history of MuSO in relation to ARC, the Europeana test case, and future steps for the continued development of MuSO. By facilitating the discovery of digital music content, and providing a virtual environment for music researchers, MuSO will promote data reuse, strengthen community standards in music representation, and create possibilities for cross-disciplinary exchange. We propose that by leveraging the connections between digital music resources and digital humanities research technologies, MuSO will facilitate new research that expands the musicological discipline.

Introduction
This paper describes the work to date on Music Scholarship Online (MuSO), an online research environment for digitized resources and digital scholarly outputs relating to music that inscribes itself within the federated model of the Advanced Research Consortium (ARC). With the project now in its third year, MuSO has reached an inflection point where it has developed a music-centered RDF schema and demonstrated the potential for federated searching across ARC nodes by crosswalking eighteenth-century music content from Europeana into ARC. The case study presented here outlines the dissemination role that MuSO proposes to play within the music research community, the history of MuSO in relation to ARC, the Europeana test case, and future steps for the continued development of MuSO. By facilitating the discovery of curated digital music content, and providing a virtual environment for music researchers, MuSO will promote data reuse, strengthen community standards in music representation, and create possibilities for cross-disciplinary exchange. We propose that by leveraging the connections between digital music resources and digital humanities research technologies, MuSO will facilitate new research that expands the musicological discipline.

Background
Researchers of music have come to depend on digital tools and information resources in their work, similar to scholars in other humanistic disciplines. Digital indexes and databases such as the Digital Image Archive of Medieval Music (DIAMM), Chopin’s First Editions Online (CFEO), and the British Library’s Early Music Online have radically simplified the task of finding scholarly resources.[1] Meanwhile, this ever-growing corpus of digitized primary source material has transformed an erstwhile preoccupation of scarcity into what is at least an appearance of abundance, extending on Rosenzweig’s remarks on born-digital records [Rosenzweig 2003]. Music researchers must now stay current with a
proliferation of new online resources to ensure that they overlook nothing of significance to their subdiscipline [Hope 2014]. They increasingly prefer digital tools and resources over print and other physical formats, even while voicing concerns over incompleteness and the superficiality of working with digitized materials [Inskip and Wiering 2015]. There is, moreover, a consensus that digital technologies have had, and continue to have, a transformative effect on scholarly networks and the work of interpretation. The ease and speed of access to digital research inputs and outputs, and the shifts in methodological scope from close to distant reading, though not yet broadly shared in musicological circles, open significant new research prospects [Pugin 2015] [Kent-Muller 2017] [Urberg 2018].

Even so, systems of scholarly production, review, and dissemination have not fully adapted to the digital realm. The infrastructures of print and publication that have dominated musicological dissemination continue to shape the discipline in the digital domain, both in the content that is produced and in the formats of that content. Traditional ideas about what is worthy of study have influenced digitization decisions and consequently the digital research that is now being undertaken on those resources. If, as Hooper notes, the allowable topics of musicological enquiry have diversified over the past few decades, the digital offers an even greater opportunity for researching underrepresented areas in musicological canons [Hooper 2016]. As Johnson has argued, these marginalized voices are valuable, acting as the “blind spots on the map, the dark continents of error and prejudice” that “carry their own mystery” [Johnson 2006]. Without the compulsion to sell enough printed stock to offset the costs of producing that stock, digital scholarship offers, at least in theory, an opportunity to further democratize music history, contrasting well-known exemplars with lesser-known, non-canonical voices. But in an academy in which recognition and reward remain tied closely to the published word, and therefore to the pressures of producing content that aligns with established cultural preferences, digital recovery work is noticeably scarce [Duguid 2014].

Beyond promoting existing content biases, the majority of digital publishing has also simply shifted formats from an analog monograph or journal article to an epub or pdf. The rich multimedia capabilities of the digital are therefore largely underutilized in a music scholarship whose digital outputs still tend towards the static, text-centric restraints of the printed page. To be fair, there are digital projects in music that do display virtuosic technical accomplishment, but the content of these projects rarely receives the same level of theoretical attention.

The traditionalist skew of digital scholarly media belies the fact that the use of digital tools across the musical domain is complex and overlapping. The boundaries between the study of musical texts (scores) on the one hand and the study of musical sounds (performances) on the other, let alone the various aims supported by digital technologies such as training, analysis, dissemination, and music making, are persistent even while fuzzy [Duffy 2009]. In addition, the radical potential of digital technologies to unify music scholarship through shared resources and methodologies remains unrealized. Siloed projects, isolated research communities, and solitary work cultures have constrained the impact of such technologies to a relative few.

This mirrors the state of literary scholarship over ten years ago when Jerome McGann and Bethany Nowviskie produced a founding document for the Networked Infrastructure for Nineteenth-Century Electronic Scholarship (NINES). In it, they described the state of the field of aggregation for digital humanities scholarship: “what you see now on the web is what you get: an agglomeration of sites and projects whose content is atomized and whose scholarly and educational value is indeterminate” [Nowviskie and McGann 2005]. NINES was founded to “explore the informational design for scholarly work at a global scale;” the goal was to develop a structured online environment that would “promote access and repurposing” of digital scholarly outputs [McGann 2011]. For NINES, scholar-produced digital projects included early digital humanities efforts such as The William Blake Archive and The Rossetti Archive.

A musical solution

This “atomized” digital content, first described by McGann and Nowviskie within a literary context, is becoming more prevalent within music scholarship. Resources such as The Lost Voices Project or Digital DuChemin, Hearing Wagner, and Songs of the Victorians are examples of the ways interdisciplinary and multimodal research outputs can utilize and extend content available through digital library and archives. The valuable research materials offered by these projects, which include biophysical data, newly accessible digitized and annotated scores, and dynamic digital editions

and visualization tools, are not readily findable to those scholars because their status as self-published scholarly outputs precludes their inclusion among many library- and publication-based digital catalogs and search engines.

Music Scholarship Online (henceforth MuSO) is working to improve the dissemination of music scholarship by making digital scholarly outputs discoverable alongside digitized music resources. With the term digital scholarly output, we refer to content that was created primarily for digital dissemination, denoting a shift away from a unidimensional print paradigm. This content may include a variety of data, media, and formats, but at its core it is content that has been transformed intellectually through the actions of experts. To give an example, Freischütz Digital is a digital edition of the opera Der Freischütz. Not only does this resource provide access to digitized manuscripts and printed editions of the opera and its libretto, it offers multitrack audio recordings, XML-based digital editions inclusive of editorial interventions made by Carl Maria von Weber, Friedrich Kind, and various copyists working with the creators, as well as the researchers’ analytic commentary that extends upon the data provided by these sources. On occasion, digital scholarly outputs include software tools, datasets, and computer code, in addition to a web presentation for the general public. The Centre for the History and Analysis of Recorded Music (CHARM), for instance, provides a search interface for its sound files as well as a suite of software tools for conducting analyses of recorded sound. Similar to MuSO in its aims, the Virtual Library of Musicology (ViFaMusik) is an information resource that provides access to a range of materials that includes primary and secondary resources. It also promotes networking opportunities with a database of music scholars, which is at this time oriented towards the German-speaking musicological world [Hope 2014] [Platt 2013]. Moreover, although ViFaMusik lists digital scholarly outputs like Freischütz Digital, these resources are not treated with the same level of granularity as the digitized content. For instance, one cannot search for a particular musical edition or digital object made available by the Freischütz Digital project. MuSO seeks to expose the constituent parts of digital scholarly outputs, making them as discoverable as digitized resources, and thus to maximize the potential for large-scale digital research.

MuSO began in 2015 as a Digital Humanities Start-up Grant funded by the National Endowment for the Humanities. At the initial planning stage, music librarians, music encoders, and musicologists gathered to discuss issues surrounding aggregation and peer review for digital scholarly outputs in music [Duguid 2016]. During this meeting, the group decided to follow in the footsteps of McGann and Nowviskie’s NINES by becoming a member of the Advanced Research Consortium (ARC). As a consequence, MuSO will build on ARC’s expertise in curating digital scholarly outputs alongside digitized collections. Through advisory and peer review activities, MuSO will identify and evaluate specialized music resources, and aggregate selected resources together with any associated data and software. In doing so, MuSO will strengthen community standards in the representation of music data and promote data reuse. Through federated searching across ARC nodes, MuSO’s users will be able to discover high-quality, vetted scholarly resources that have been curated by experts in other humanistic disciplines, and thereby conduct thorough, multidisciplinary research. With these efforts, MuSO will advocate for open, collaborative, and cross-disciplinary research practices in music.

ARC: A collaborative for digital research in the humanities

ARC is a federation of scholarly organizations that identify, review, and curate digital content most relevant to the community of scholars each organization serves, e.g. NINES serves nineteenth-century scholars and 18thConnect, eighteenth-century scholars. Socially, the ARC organization is made up of representatives from digital research environments like NINES and MuSO. Technologically, ARC is the home of the “ARC Index” — the physical computing and software infrastructure that aggregates reviewed and curated digital content into one single catalog that then provides high-quality scholarly resources for humanities scholars.

ARC is not a digital repository (like Bepress, DSpace, or Fedora Commons) or a publishing platform (such as the Manifold digital publishing platform[6]), nor is ARC a digital library environment, like the Perseus Digital Library or HathiTrust Digital Library, as ARC does not host or publish digital content. This means that, instead of hosting resources, such as digitized page images, scholarly presentations and papers, or book publications, ARC ingests metadata about digital resources that already exist on the web and have been reviewed or curated by ARC’s scholarly organizations.
ARC’s research environments are digital aggregators, and each research environment prescribes a set of peer review guidelines for the inclusion of a resource. Usually, scholar-created digital projects undergo blind peer-review and databases of digitized content undergo community approval processes. When a digital resource, project, or database has been approved, metadata about the “holdings” (or discrete digital items) is ingested into the “ARC index.” Once ingested into the index, this metadata allows each research environment to provide a search experience that aggregates records describing respected, scholarly digital resources into one federated search and discovery system. Search is open, whether across ARC nodes, or within individual ARC research environments, but each ARC node may choose to aggregate content that is behind a paywall according to its prescribed guidelines or review policies.

When a scholar searches the ARC index, the experience is similar to searching within an institution’s library search interface or the Digital Public Library of America, in which searches populate a list of results that point to a physical (in the stacks of a library) or digital space (in a database like JSTOR) where the item lives. Like DPLA and library search interfaces, using ARC for research purposes will return a list of results that point to where digital content can be located. Unlike other scholarly research portals online, ARC also points scholars to curated digital scholarly outputs created by and relevant to humanities researchers. These digital scholarly outputs, which range from annotated digital editions to complex archives, are indexed within ARC, and the digital content findable within ARC’s catalog contains not only a single metadata record pointing to the project, but many records describing the content available within the project.

ARC’s digital aggregation efforts have resulted in a catalog of over two million curated digital objects, spanning the medieval period and up to copyright, in federated digital research environments such as those listed on the right-hand side of Figure 1. The resulting federated ARC network allows scholars to search for content in previously “atomized” scholar-produced projects. Once spread far and wide across the internet, the digital projects aggregated by ARC are now discoverable via a single, scholarly, and scholar-designed search interface. 

![Figure 1](image_url)
ARC intends to reflect the interdisciplinary nature of digital humanities. Since digital work is often collaborative and interdisciplinary, an aggregated body of digital research inputs and outputs should consequently be findable and accessible to scholars from various disciplines. For digital methods to be fully integrated as humanities research methodologies, the outputs of these methods must be familiar to all humanists: all of the multidisciplinary, multilingual and international scholars that our institutions serve. Against the backdrop of this rich history of interdisciplinary and non-hierarchical work, MuSO joins ARC as a research community that seeks to build on ARC’s engagement with humanities scholars, by serving music scholars.

MuSO and Europeana

At the conclusion of the NEH start-up grant, the MuSO team recognized the need for a new way to describe digital objects in music that would bring digital scholarly outputs in music together with digitized resources. At the same time, the schema and its descriptive ontologies needed to be accurate, while still broad enough to ensure interoperability across the boundaries of discipline and format within ARC. This discovery-level schema would not be intended to supplant the valuable, highly descriptive preservation-level data being generated by libraries, as MuSO would still send users to the resource and its accompanying preservation-level metadata to conduct more in-depth analysis. Rather, MuSO’s metadata schema would need to draw out the most important pieces of information that researchers need for discovery. To help with the creation of such a schema, MuSO chose to rely on yet another digital aggregator, Europeana.

Europeana “is the organisation tasked by the European Commission with developing a digital cultural heritage platform for Europe” [Kenny 2017]. Since its launch in 2008, Europeana has been working with libraries, museums, and other organizations in the cultural heritage sector throughout Europe to aggregate their digital collections into a single catalog and online interface. In the autumn of 2016, Europeana instituted a new grant scheme, entitled the Europeana Research Grant scheme, for scholars to work with its aggregated collections. Building on Europeana’s expertise in standardizing metadata descriptions of widely disparate collections, this grant would allow MuSO to construct an initial metadata schema, which could bridge the gap between Europeana’s metadata schema, the Europeana Data Model (EDM), and the ARC RDF schema while maintaining and incorporating the long-established cataloging standards of the music research community.

The new MuSO schema would allow music scholars to use MuSO for its music-specific content, but they could then expand their search to encompass the period-specific offerings of 18thConnect, the node of the Advanced Research Consortium (ARC) dedicated to eighteenth-century studies. MuSO’s federated searching capability therefore would become an opportunity to “widen and bridge research fields” as well as to promote knowledge of digital humanities research tools and methodologies within the musicological community [Pugin 2015]. In other words, MuSO would provide users with what they are increasingly demanding: the ability to isolate content relating to Haydn’s “The Flowers of Edinburgh,” Hob. XXXIa:90 (for instance), regardless of whether it is in the context of a website covering eighteenth-century folk songs, Scottish literature, or a complete Haydn digital edition.

Aggregating Eighteenth-Century Material from Europeana Music

In the first phase of the Europeana Grant, the team, consisting of a subset of the original MuSO team, inspected the metadata standards of over a dozen other projects and organizations, including the Digital Image Archive of Medieval Music (DIAMM), Répertoire International de Littérature Musicale (RILM), Répertoire International des Sources Musicales (RISM), Répertoire international de la presse musicale (RIPM), the Library of Congress’ Music Treasures Consortium, and others in order to identify common traits which might inform our decisions, as well as more customized metadata requirements which might also be useful. The overarching categories identified from these and other metadata standards included repository identifier, item identifier, language, data, physical description, title, and statements of responsibility.

With this information, the team entered the prototyping phase of the project. The team met for two days to discuss and formulate an initial metadata schema based on the information that was gathered in phase one. We distinguished where
existing standards such as Dublin Core provided adequate structure, and where it might be necessary to propose specialized fields.

The third phase of development was necessarily iterative, as the team's abstract model made contact with actual data. During this phase, the team "crosswalked" eighteenth-century content in the Europeana Music Collection over to ARC using the new MuSO schema. A crosswalk is designed to show users or database managers how to translate data from one descriptive formatting to another. Because multiple metadata schemas exist in the humanities and cultural heritage sector, e.g. Dublin Core, MARC, TEI, RDF, such an intellectual and technical mapping must take place to allow search and indexing technologies to speak to each other. In general, the workflow followed a classic Deming cycle: Plan, Do, Study, Act. In general, the workflow followed a classic Deming cycle: Plan, Do, Study, Act. The workflow was progressively refined as implementation produced new information to incorporate: some elements which seemed desirable in the planning stages were not practical to implement in this phase or were not germane to the actual content of the data.

For instance, one element that seemed highly desirable was notation form, in order to distinguish systems like Gregorian chant notation, and lute tablature from modern standard notation. However, retrieving such information would have required item-level examination, as notation was inconsistently described in the dc:description field. And while the Resource Description and Access vocabulary for notation type was integrated into the MuSO metadata schema, the team chose to defer full notational description in view of time considerations.[8]

The team had also hoped to use the Library of Congress Genre/Form Project vocabularies in its subgenre field for maximum interoperability, but that project is still in progress and has not yet published an official vocabulary. In its absence, the team carried over Europeana fields that contained information on musical form and genre. This information will be retained until a standardized vocabulary is established.

Variations and areas for interpretation inevitably arose, even in a standardized structure such as Europeana's. One of these areas was an unforeseen complication: institutions with multiple RISM sigla. RISM sigla have become the standard in the music community for identifying cultural heritage institutions (libraries, archives, museums, etc.) that hold music source materials. These sigla consist of a two-letter country code followed by a unique 3-4 letter institution designation.[9] Though the sigla are intended to be unique, some institutions have become consolidated since their sigla were assigned. This has led to many institutions such as the Bibliothèque nationale de France having several sigla assigned to them. The MuSO team therefore had to determine which sigla to use. In another example of the data variations in the Europeana dataset, certain art forms required consensus due to their multidisciplinary nature, such as ballet, which incorporates music, dance, and drama. Some cases also pointed to a clear need for new genre terms in the ARC vocabularies, which the team suggested: e.g. dance, pedagogy, ethnography, and sculpture/architecture.

Similarly, a consensus was necessary on the English translation and labeling of terms describing roles and responsibilities, as with Éditeur scientifique and Compositeur prétendu.[10] This process was also iterative, as team members conferred and reconciled interpretations and decisions across datasets. Retrospective editing was not difficult, as the facet and clustering function in OpenRefine, which is an open source tool for working with messy data, aided finding and replacing terms as needed.

With the data successfully crosswalked, the fourth phase focused on ingesting the data into 18thConnect. During this time, the team worked with the development team at 18thConnect to ensure that the data was consistent and that it displayed properly on the 18thConnect website. Furthermore, the team launched a new MuSO website that would highlight MuSO's schema and aggregated content.[11]

**Looking Forward**

Despite the advances of the MuSO project, much work remains in generating a digital portal that is fully capable of curating scholarly research and materials. During the phases of the Europeana grant and the initial Start-up Grant from the National Endowment for the Humanities, the MuSO community of scholars identified four significant tasks for the
Thanks to the efforts of 18thConnect, MuSO has been able to make a selection of the content found in Europeana Music interoperable with the existing ARC catalog. This is an important step, as it allows scholars to discover articles about Mozart alongside the actual letters and scores that those articles cite. Nevertheless, ARC’s reliance on text-based metadata means that its current search interfaces are completely biased towards the search and discovery of textual information, even if that information is describing some non-textual content. Although music can be notated and encoded in ways similar to text, it is also heavily reliant on sonic and non-textual manifestations, and a number of researchers are working on these issues of musical search for both audio and symbolic formats. Given MuSO’s commitment to interdisciplinarity, its search interface must bring these efforts in non-textual search together with existing textual search technologies, expanding ARC’s search capabilities to allow scholars to search and discover melody, harmony, and rhythm in addition to text.

MuSO’s collaboration with the ARC community will likely result in more than an expansion of search capabilities. ARC was originally organized according to temporal boundaries such as nineteenth-century studies, medieval studies, and modernist studies. MuSO is one of the latest in a number of recent additions to ARC that is dedicated to a particular topic or discipline. MuSO’s efforts to promote digital scholarly outputs alongside digitized musical resources are spurring the ARC community to re-evaluate its curatorial practices. It is not enough to append a new music-based research environment to ARC’s current structure. MuSO will invariably aggregate content of relevance to ARC’s other research communities (i.e. Wagner-related content for nineteenth-century scholars). MuSO’s future efforts will therefore include partnering with the existing research communities in ARC to ensure that digitized musical resources are discoverable in the relevant research environments outside of the MuSO portal so that Wagner-related content can be discovered in NINES along with MuSO. Moreover, MuSO will work with the ARC community to develop methods of conducting both period-specific and discipline-specific review of digital scholarly outputs that will provide relevant peer-review for contributors while assuring users that the content made available through ARC and its nodes has been vetted by experts in relevant disciplinary and cultural-temporal fields.

MuSO will continue to identify and aggregate existing digital collections into its catalog. Resources such as Europeana, HathiTrust, and the DPLA continue to grow, and MuSO will do the same, identifying digitized primary sources and secondary scholarship for inclusion in the MuSO catalog. MuSO will also work to include digital scholarly outputs in its catalog. It has already identified a number of digital projects that meet technical and scholarly standards that would likely pass peer review and could be incorporated into the MuSO catalog. MuSO will reach out to these projects and ones of similar quality that arise in the future, encouraging them to consider digital peer review.

This goal is closely related to MuSO’s final future goal to promote digital outputs and metadata standards among music scholars. MuSO will work closely with digital projects to ensure they follow best practices for describing their digital outputs, and it will promote those outputs by making them available for the scholarly community to search and discover alongside other relevant scholarly resources. The criteria for inclusion in MuSO remain under development in anticipation of the continued evolution of digital scholarship in music. In addition to accommodating emerging technologies and methods, our intention to democratize participation in digital scholarship favors an inclusive approach to proposed projects. Therefore thinking “outside the box” may find a place as “a feature, not a bug,” where traditional modes of scholarly communication in music research retain a formidable pedigree. The basic criteria for inclusion established in the January 2016 meeting for MuSO continue to apply: Does the project make a worthwhile contribution to research in its subject area? Is its methodology sound? Does the project achieve its own goals? Are there major obstacles to usability? Is its existence and accessibility sustainable? [Duguid 2016]

Scholars are increasingly turning to digital resources for conducting and disseminating their research. Professional societies such as the Modern Language Association [MLA 2017] and the American Historical Association [AHA 2015] are setting standards for digital outputs, and governmental standards initiatives such as the United Kingdom’s Research
Excellence Framework are making allowances for digital outputs. However, current outlets for reviewing music scholarship are optimized for static print and therefore struggle to assess multi-modal and multidisciplinary digital scholarly outputs. Amidst a growing set of high-quality digital projects, the musical community demands not only a set of standards for assessing born-digital scholarship but it demands a community with expertise in assessing that scholarship [Duguid 2014]. Without a central location afforded by traditional publishing platforms such as journal articles and monographic series, digital scholarly outputs demand digital aggregation platforms that ensure they are discoverable by the scholarly community. MuSO is poised to step into this void, promoting quality digital scholarship in music while giving it an equal voice alongside the traditional platforms of dissemination.

Notes

[1] For DIAMM, see https://www.diamm.ac.uk/; for CFEO, see http://www.chopinonline.ac.uk/cfeo/; for Early Music Online, see https://www.royalholloway.ac.uk/music/research/earlymusiconline/home.aspx.


[10] Éditeur scientifique ordinarily designates an editor of a scholarly monographic series or journal, however in the case of the eighteenth-century score anthologies held in the Bibliothèque nationale de France, the term was used to describe the role of the person responsible for compiling and editing the songs published therein. Consequently, the MuSO team agreed to use the “Compiler” relator code for this term. Compositeur prétendu described a circumstance in which a work’s creative attribution was later called into question; the MuSO team preserved this term in its metadata as “Attributed name.”


Works Cited


