Interinstitutional Thematic Collection Development: Technical and Procedural Considerations from the Sounding Spirit Digital Library

White Paper

By Jesse P. Karlsberg and Erin Fulton

July 2020


Project Director: Jesse P. Karlsberg

Sounding Spirit (https://soundingspirit.org) is an initiative of the Emory Center for Digital Scholarship, Emory University, Atlanta, Georgia.
Interinstitutional Thematic Collection Development: Technical and Procedural Considerations from the Sounding Spirit Digital Library

White Paper

By Jesse P. Karlsberg and Erin Fulton

July 2020


Project Director: Jesse P. Karlsberg

Sounding Spirit (https://soundingspirit.org) is an initiative of the Emory Center for Digital Scholarship, Emory University, Atlanta, Georgia.
# Table of Contents

Summary .................................................................................................................................................. 5
Audience .................................................................................................................................................. 5
Recommendations in Brief ...................................................................................................................... 5

Processes and Outcomes: Sounding Spirit as a Case Study .............................................................. 6
Music Bibliography ............................................................................................................................... 6
   Context .................................................................................................................................................. 6
   Process ................................................................................................................................................ 7
   Outcomes ............................................................................................................................................ 8
Partnerships/Collaboration .................................................................................................................... 8
   Context ................................................................................................................................................ 8
   Process ............................................................................................................................................... 9
   Outcomes ........................................................................................................................................... 9
Technical Workflows ............................................................................................................................. 10
   Context ............................................................................................................................................... 10
   Process ............................................................................................................................................... 11
   Outcomes ........................................................................................................................................... 13
Pilot Site .................................................................................................................................................. 14
   Context ............................................................................................................................................... 14
   Process ............................................................................................................................................... 14
   Outcomes ........................................................................................................................................... 15
Future Phases .......................................................................................................................................... 15
   Context ............................................................................................................................................... 15
   Process ............................................................................................................................................... 16
   Outcomes ........................................................................................................................................... 16

Recommendations in Detail .................................................................................................................. 17
Build Digital Thematic Research and Teaching Collections .............................................................. 17
   Why? .................................................................................................................................................. 17
   How? ................................................................................................................................................. 18
Foster Collaboration in Digital Collection Development ...................................................................... 19
   Why? .................................................................................................................................................. 19
   How? ................................................................................................................................................. 19
Accommodate Institutions with Diverse Affordances in Digitization and OCR Workflows ............... 20
   Why? .................................................................................................................................................. 20
   How? ................................................................................................................................................. 20
# Appendix A: Prioritized Criteria and Search Processes

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prioritized Criteria</td>
<td>22</td>
</tr>
<tr>
<td>Search Process</td>
<td>23</td>
</tr>
</tbody>
</table>

# Appendix B: Metadata Schema

Page 28

# Appendix C: Digitization Processes Questionnaire

Page 33

# Appendix D: Memorandum of Understanding Template

Page 34

# Appendix E: Digital Library Use Case

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>37</td>
</tr>
<tr>
<td>Image Quality</td>
<td>37</td>
</tr>
<tr>
<td>Federated Collections</td>
<td>38</td>
</tr>
<tr>
<td>Selectable Text</td>
<td>38</td>
</tr>
</tbody>
</table>

# Appendix F: Digitization Workflow

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope</td>
<td>39</td>
</tr>
<tr>
<td>Roles</td>
<td>39</td>
</tr>
<tr>
<td>Tasks</td>
<td>39</td>
</tr>
<tr>
<td>Sample Scan Review Rubric</td>
<td>41</td>
</tr>
</tbody>
</table>

# Appendix G: Visualizations of List of Digitization Candidates

<table>
<thead>
<tr>
<th>Visualization</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of Publication by Decade</td>
<td>43</td>
</tr>
<tr>
<td>Place of Publication and other Associated Places</td>
<td>44</td>
</tr>
<tr>
<td>Music Notation Systems</td>
<td>45</td>
</tr>
<tr>
<td>Non-English Languages</td>
<td>46</td>
</tr>
<tr>
<td>Denominational Affiliations</td>
<td>46</td>
</tr>
<tr>
<td>Characteristics Represented</td>
<td>47</td>
</tr>
</tbody>
</table>

# Appendix H: Lists of Descriptive Entries

<table>
<thead>
<tr>
<th>List of Descriptive Entries</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary Collection List</td>
<td>48</td>
</tr>
<tr>
<td>Preliminary Bibliographic Essay Topics</td>
<td>50</td>
</tr>
</tbody>
</table>
Summary

This white paper shares insights on the development of digital thematic teaching and research collections. The report draws on the yearlong planning process for the Sounding Spirit Digital Library, an open access digital resource collecting southern sacred vernacular music books published between 1850 and 1925. Funded by a Foundations grant from the Humanities Collections and Reference Resources program of the National Endowment for the Humanities (NEH), this planning process was led by a team based at Emory University’s Center for Digital Scholarship (ECDS) and staff at four archives with strong holdings of southern vernacular sacred music, with the support of an engaged advisory board.

This white paper first outlines the major activities of the planning process, which included 1) music bibliographic research informing digitization plans, resulting in the publication of a “Checklist of Southern Sacred Music Imprints, 1850–1925”; 2) the formalization of the partnership between ECDS and participating archives; 3) workshopping of interinstitutional digitization and optical character recognition processes, and the application of these processes through the digitization of twenty-two songbooks; 4) the launch of a pilot digital library site that organizes these music books into collections and enhances them with descriptive entries; and 5) the development of a plan for implementation, resulting in the submission of an application for support from the NEH Humanities Collections and Reference Resources program. For each of these components of the planning process, the white paper offers context that informed our approach, an outline of our processes, and a discussion of major outcomes.

The white paper concludes with recommendations building on our experience. These recommendations are organized around three themes: 1) thematic collection development, 2) collaboration in developing digital resources, and 3) approaches to digitization and optical character recognition.

Audience

This white paper is written with specific audiences in mind. The report is chiefly addressed to humanities scholars who work with primary sources and to library and archive staff who steward humanities collections. Intended audiences also include the administrators of scholarly societies whose members research and teach with humanities materials, funders who support the development of thematic collections, and digital humanities professionals who develop software supporting engagement with primary sources.

Recommendations in Brief

Archives engaged in digitization projects should prioritize building thematic research and teaching collections to encourage engagement with their holdings. Scholars should participate in the development of thematic collections to further the impact of their work.

- Thematic learning digital collections of primary sources intended to support research and teaching provide an avenue for increasing public engagement with humanities materials;
- Thematically organized digital resources help realize the potential of digitization for enhancing teaching and research with primary sources.
Archives should collaborate with each other, with humanities scholars, and with digital humanities centers in digital collection development.

- Digital thematic research and teaching collections will be most successful when they emerge from collaborations among archives with complementary collection strengths and among humanities scholars engaged in research with primary sources;
- Collaborative digital resource development will also benefit from the inclusion of technologists with expertise in digitization workflows and interfaces for engaging with digitized texts.

Digitization initiatives should value consistency and quality as well as the inclusion of lower-resourced institutions when developing digitization and optical character recognition (OCR) workflows.

- Digitization initiatives should include institutions with diverse affordances. Digitization projects could incorporate flexibility within defined ranges around some specifications, or partner with vendors or institutions with robust digitization programs to facilitate the participation of lower-resourced institutions;
- Including OCR that generates textual and positional information in digitization workflows can make digital collections more accessible and useful.

**Processes and Outcomes: Sounding Spirit as a Case Study**

**Music Bibliography**

**Context**

Texts were critical to the practice of vernacular southern sacred music cultures of the late nineteenth and early twentieth centuries. Yet scholarship and popular understanding of genres such as gospel, spirituals, shape-note singing, and lined-out hymn singing typically minimize the role of music books. Sounding Spirit—a research lab and publishing initiative promoting engagement with historical sacred American music books through a range of activities including the publication of scholarly editions—is developing the Sounding Spirit Digital Library motivated by the conviction that digitizing and making available large representative collections of these works will stimulate research and teaching, and contribute to scholarly and public reappraisals of these genres and their place in American history and culture.

Collections of sacred song associated with southern sacred music genres were produced in a wide variety of print formats suited to diverse circumstances of use. These works served as a central point of dissemination for repertoire and record dynamic expressions of cultural exchange across race, class, gender, and geography. Although numerous southern sacred music books achieved print runs in the hundreds of thousands, the vernacular genres that they represent were undervalued by academics at the time they were first produced. Archives concomitantly failed to prioritize collecting books that exemplified such genres, despite their great number and popularity. In the mid-twentieth century, these musics were incorporated into the American musical canon by folklorists, who misapprehended them as
primarily oral traditions, driven in part by disciplinary blinders but also by the paucity of relevant volumes in academic repositories.

Sounding Spirit aims to identify and make accessible such long-undercollected texts through digitization and thematic collection development and to highlight the centrality of print culture to sacred music practice. Because no preexisting bibliographies offered control over southern vernacular sacred music of the late nineteenth and early twentieth centuries, this project necessarily included an attempt to apprehend the extent of that corpus, guided by research questions surrounding the role of text in southern sacred music. An increasingly broad and nuanced understanding of the extant repertoire will support the development of a digital library that represents the variety of the corpus. Bibliographic research helps surface underappreciated works that might contribute to research and public understanding, complicating received narratives of vernacular sacred music genres and their role in American history and culture.

Bibliographic research also supports project goals of organizing digitized books into thematic collections of volumes that correspond to major areas of scholarly inquiry and the interests of more general readerships. The inclusion of robust descriptive metadata, drawing on bibliographic research, will facilitate interoperability with other digital repositories and improve the usability of the Sounding Spirit Digital Library.

Process
Bibliographic research entailed refining the scope of the project into a set of prioritized criteria, instrumentalizing these criteria, and developing a metadata schema to describe identified volumes.

1. Collection criteria:
   a. Developed a set of ten collection criteria drawing on project scope and motivating research questions;
   b. Prioritized criteria from greatest importance to least: year of publication, geographic origin, sacred content, vernacular content, presence of copies in partner archives, race/ethnicity, genre, and language, influence/significance, current availability;
   c. Reviewed criteria with the advisory board, both individually and in group meetings.

2. Generation of bibliography:
   a. Developed an extensive list of search terms associated with the prioritized collection criteria and indexable in a majority of library catalogs (see Appendix A);
   b. Thoroughly searched the catalogs of the four partner institutions, making modifications to search terms as necessary to intersect with local cataloging practice;
   c. Supplemented this search with additional searches of selected non-partner institutions and the Online Computer Library Center (OCLC) union catalog to better encompass audiences, genres, periods, and geographies underrepresented in the holdings of the partner institutions;
   d. Transcribed bibliographic data from each relevant catalog record, along with project-specific information related to the collection criteria, into a comprehensive checklist.

3. Metadata schema:
   a. Collected standard bibliographic data corresponding to the Dublin Core schema for each relevant catalog record;
   b. Designated further metadata fields to facilitate interoperability, particularly those required by the Digital Library of Georgia (DLG)/Digital Public Library of America (DPLA);
(c) Identified additional considerations specific to this corpus that could usefully be included in the descriptive metadata, such as musical notation type, denominational affiliation, width (in addition to height) of volumes, and associated geographical locations other than the place of publication;

(d) Developed draft metadata schema for a future implementation phase including further project-specific metadata points drawing on characteristics collected during the compilation of the checklist.

Outcomes

Sounding Spirit’s music bibliographic research resulted in a “Checklist of Southern Sacred Music Imprints, 1850–1925” (https://doi.org/10.15139/S3/VQT4D5) and a draft metadata schema (see Appendix B). Bibliographic research led by Sounding Spirit music bibliography associate Erin Fulton, with the assistance of project director Jesse P. Karlsberg, yielded an initial list of 2,265 relevant volumes. Sounding Spirit published the checklist in tab-delimited format in the open access Sounding Spirit Dataverse hosted by the University of North Carolina’s Odum Institute. The checklist assembles basic bibliographic data otherwise scattered across the catalogs of numerous repositories. The dataset also includes thirty-three data points that standard catalog records do not encompass, indicating whether each work meets criteria around audience, genre, context, or style. A detailed data dictionary accompanying the checklist defines each field of the checklist, explains how to interpret its value for a given work, and indicates the sources for that information.

The vernacular sacred songbooks represented in this collection are significant markers of the influence and creativity of a wide range of underrepresented populations not yet fully acknowledged as crucial contributors to American history and culture. The checklist represents the first attempt at bibliographic control over this rich and varied repertoire. These music books document little-known publishing practices such as the hymnals and missionizing texts of regional denominations; the words-only hymnals used by displaced Creek, Cherokee, and Choctaw populations; the prolific songwriting of black Pentecostal Christians; and the musical activities of Polish, Czech, Greek, Dutch, and German immigrant populations. The project team relied on this initial checklist in developing a smaller selection of candidate volumes for digitization, discussed below under Future Phases.

The planned metadata schema will facilitate discoverability of digital library holdings, both within the federated collection and via other platforms such as the DPLA and the OCLC. The inclusion of rich descriptive metadata, including data points specific to this project, will permit the project team to offer additional pathways for access and engagement through the creation of interactive visualizations such as timelines and maps. Compliance with external metadata standards, including Dublin Core, ensures interoperability with other online repositories.

Partnerships/Collaboration

Context

Sounding Spirit’s project team, based at Emory University’s Center for Digital Scholarship (ECDS) partnered with four archives due to their strong holdings of southern vernacular sacred music: Pitts Theology Library and Stuart A. Rose Manuscript, Archives, and Rare Book Library at Emory University; the
Center for Popular Music at Middle Tennessee State University; the John Jacob Niles Center for American Music at the University of Kentucky; and the Archives and Special Collections of Southern Baptist Theology Seminary. Each contributing institution offers unusually rich holdings within the historically under-collected subject area of American vernacular sacred song. In its English and American Hymnody and Psalmody Collection, Pitts Theology Library offers the second-largest collection of hymnals and psalters in North America. Emory’s Stuart A. Rose Manuscript, Archives, and Rare Book Library includes numerous sacred songbooks within its African American History, Rare Books, and Southern History collections. Established in 1985, the Center for Popular Music at Middle Tennessee State University has grown to become one of the nation’s largest American music research centers, holding over one million items. The John Jacob Niles Center for American Music at the University of Kentucky provides a nexus for the research and performance of American music, with special emphasis on cultures of the southeastern United States; central to its rare book holdings is the 8,000-volume collection of tunebooks and hymnals originally assembled by sacred music scholar Glenn C. Wilcox. Founded in 1859, the Southern Baptist Theological Seminary holds a collection of over 5,000 hymnals and songbooks, including significant representation of gospel and Sunday school songbooks, collections of spirituals, and shape-note tunebooks.

Because of their locations near southern centers of music publication and performance, institutional relationships with collectors and practitioners of sacred music, and missions that emphasize the study and practice of sacred and/or American music, these four archives offer a robust cross-section of the corpus of volumes identified through bibliographic research. Moreover, each institution varies in organizational structure and policy; availability of digitization equipment and staff; capabilities and established practices for post-processing, including OCR; and preservation and access practices. These differences facilitated the development of extensible workflows that can serve as models for other collaborations among institutions of varying sizes and affordances.

**Process**

(1) Distributed a digitization processes questionnaire (see Appendix C) to partner institutions, soliciting information on their capabilities, policies, and practices;
(2) Solicited sample scans from each institution;
(3) Conducted quality assurance to test the digitization apparatuses employed in the production of those scans (see below under Technical Workflows);
(4) Negotiated and signed memoranda of understanding with each institution, outlining its respective responsibilities during the pilot project and indicating commitment to pursue implementation;
(5) Received digitized sample volumes from each institution for inclusion in the pilot digital library (see below under Pilot Site);
(6) Prepared for institutional participation in the implementation phase, including vetting of staff, responsibilities, and scope of effort for each partner archive.

**Outcomes**

Each partner institution signed a memorandum of understanding (see Appendix D). These MOUs outline the relationship between ECDS and the partner institutions for both the planning and implementation phases of the Sounding Spirit Digital Library project. ECDS has assumed responsibility for developing the digitization and OCR standards used in the project, hosting the Readux site, ingesting digitized volumes into Readux, and seeking funding for implementation. Partner institutions have committed to reviewing and approving a list of volumes from their collections for inclusion in the digital library; adhering to
agreed-upon standards for digitization, OCR, and preservation; and identifying needs and sources for implementation funding. Importantly, these commitments include considerations for the long-term future of the project. Partner archives have agreed not only to obtain approval for the digitization of materials selected for implementation, but also to ensure the ongoing preservation of digitized volumes in their collection.

Technical Workflows

Context

The production of the Sounding Spirit pilot digital library involved the digitization of a handful of works from each partner archive and the application of post-processing and OCR to the digitized works. Accomplishing these tasks also involved research into standards and software associated with digitization, OCR, and the automated recognition of music in the digitized works.

Digitization has long been a major means for archives to enhance access to humanities collections. Yet digitization strategies and workflows are in flux as technologies improve, use cases proliferate and diverge, and institutions develop varied capacities depending on their priorities and budgetary constraints. At many institutions, current equipment and processes associated with book digitization were put in place to meet patron requests for scans, or to support digital collection development focused on high-volume black-and-white digitization suitable for text extraction. Sounding Spirit’s emphasis on the materiality of historical music books places the project among a growing number of digitization initiatives requiring high-volume digitization and producing high-resolution, full-color images while also accommodating machine recognition of text and other information. National and international standards such as the Metamorfoze Preservation Imaging Guidelines and those of the Federal Agencies Digital Guidelines Initiative (FADGI) have facilitated the development of digitization workflows that meet the needs of projects such as Sounding Spirit. Such initiatives support the production of consistent and high-quality digital images, taking into consideration features including resolution, color space and accuracy, focus, and distortion. Yet legacy equipment purchasing patterns mean that many institutions are poorly positioned to execute such projects. Some institutions—such as Emory University—perform high-volume digitization using equipment purchased to facilitate the rapid production of tightly cropped black-and-white images. Smaller institutions that have prioritized meeting patron requests have typically purchased flatbed scanners, which may fall short of the quality standards of projects like Sounding Spirit or cannot readily accommodate efficient high-volume digitization. A range of vendors offer digitization services aimed at meeting the needs of projects and institutions that lack onsite capacity for high-quality and high-volume digitization, offering compliance with FADGI standards.

OCR is inconsistently included in archives’ digitization workflows. Optical music recognition (OMR) or other forms of music information retrieval (MIR) are seldom incorporated into digitization processes. Yet the results of OCR and MIR have the potential to dramatically enhance the usability of digital humanities collections for scholarly and public audiences. Browsing interfaces—such as that employed by Readux, the platform the Sounding Spirit Digital Library adopts—can draw on textual as well as positional information derived from OCR to support richer engagement with digitized works through features like text selection and annotation. Search tools can query OCR results, improving the discoverability of works within a digital collection. Experimental search interfaces illustrate the potential that music information
offers for enhancing access. Extracted textual and musical information also creates opportunities for researchers, including through computational methodologies.

OCR is a rapidly advancing field and OCR workflows frequently lag behind cutting-edge research. A 2018 report commissioned by NEH and the Andrew W. Mellon Foundation offering “A Research Agenda for Historical and Multilingual Optical Character Recognition” identified challenges that inhibit effective humanities research drawing on digital texts and has spurred the NEH to call for software development addressing the shortcomings of OCR.¹ Presentations at the January 2020 Mellon Workshop on OCR and Digital Text Production at the University of Maryland demonstrated that open source OCR engines such as Tesseract, Kraken, and Ocropus are rapidly improving in accuracy across languages and historical periods, and in other areas such as layout detection and read order. Commercial OCR services including Google Cloud Vision, Microsoft Computer Vision Read API, Abbyy Cloud OCR, and Amazon Textract are also rapidly improving. Yet the cutting-edge OCR offered by these technologies remains out of reach for most, particularly for smaller and lower-resourced archives or researchers who either lack the technical knowledge needed to conduct OCR on their own or the resources to pay vendors to conduct it for them. Other commercial tools may provide adequate functionality but frequently require continued costs. When institutions do conduct OCR, they frequently extract only textual information in plain text format from digitized works, neglecting to collect or retain other valuable information that OCR engines can produce such as positional data and confidence scores.

OMR and MIR are likewise rapidly developing fields. Although OMR remains inaccurate and labor-intensive, including more readily obtainable musical information in the digital library will enhance the collection’s research potential by facilitating the extraction of regions of pages that contain music.

The potential for research with digital collections that include musical and textual information alongside metadata and digitized page images points toward the intersection between humanities collections, reference resources, and documentary editions. Reference resources important to American music, such as the Hymn Tune Index and the unpublished Southern and Western American Sacred Music and Influential Sources, enable research into the publication histories of hymn tunes by facilitating searches of transcribed melodic incipits in addition to textual metadata. International collections of digitized music sources like Répertoire International des Sources Musicales similarly facilitate research through supporting queries of incipit databases. But such reference resources haven’t exhaustively encoded music. By including textual information as well as musical information where feasible, Sounding Spirit intends to acknowledge the value of offering digital representations of the works themselves as well as extracted information. This approach highlights the potential for future convergence between thematic collections and reference resources. Sounding Spirit’s digital library planning process included running OCR on digitized works, and assessed the applicability of OCR, and OMR, and other MIR techniques to its music books.

Process

(1) Developed use case to inform digitization quality standards (see Appendix E):

(a) Reviewed international guidelines on digitization best practices, including Metamorfoze and FADGI;
(b) Developed use case guided by anticipated audiences and the scholarly research questions guiding their uses of the digital library;
(c) Drafted digitization guidelines modeled on FADGI 3 and Metamorfoze Light, with additional requirements based on specifics of the materials and some allowances for differences in partner archive affordances.

(2) Reviewed sample scans:
   (a) Drawing on responses to digitization processes questionnaires, selected equipment at each partner archive best aligned with project use case and digitization guidelines;
   (b) Partner archive staff produced sample scans of works selected for digitization;
   (c) Reviewed sample scans to assess attainment of standards outlined in digitization guidelines and suggested adjustments to equipment or processes where needed.

(3) Ran OCR study:
   (a) Surveyed OCR practices at partner archives, determining that two of four partners typically produced OCR: one using Tesseract to generate plain text OCR, the other using ABBYY FineReader to produce Alto XML, including both textual and positional information;
   (b) Developed a reference corpus by transcribing text for a representative selection of sample scans of digital library works;
   (c) Processed page images associated with the reference corpus using six leading OCR engines and evaluated the results:
      (i) Ran OCR using leading open source (Kraken, Tesseract) and commercial (ABBYY Finereader SDK, Amazon Textract, Google Cloud Vision, and Microsoft Azure Cloud services) engines;
      (ii) Identified limitations and challenges impacting the applicability of our findings, including:
            1) Our reference corpus included only English-language works, though our digitization candidates include seventy-nine works in thirteen additional languages;
            2) We may have made errors transcribing the pages in our reference corpus;
            3) The two open source engines we tested require training to achieve the best results, but performing training was beyond the scope of our test.
      (iii) Compared results from each engine to transcription using ocrevalUAtion tool and identified GCV and MACV as best performing engines on corpus;
            1) Identified juxtaposition of music and text as major issue for many OCR engines, which sometimes erroneously identified music as garbled text;
            2) Identified layout analysis and reading order as issues for all OCR engines, including the best performing engines. These issues impact the quality of OCR for the Sounding Spirit corpus, which features multiple complex juxtapositions of text and music blocks.
            3) Identified technical requirements and the lack of user-friendly interfaces as issues impacting access to cutting-edge OCR engines.
      (iv) Presented on findings at January 2020 Mellon Workshop on OCR and Digital Text Production.

(4) Developed flexible digitization and post-processing workflow (see Appendix F), including:
(a) Contributing archive captures digitization target, scans volume, and performs quality
assurance checks (QA) for completeness, focus, color, and distortion;
(b) ECDS performs additional QA;
(c) Partners make any needed adjustments;
(d) ECDS performs post-processing using Adobe Creative Suite and GoldenThread software
with input from partner archives;
(e) ECDS runs OCR using GCV and creates access copies of all files.

(5) Carried out digitization and ran QA according to workflow, documenting challenges for
implementation, including:

(a) Crop challenges: ensuring that all edges of a given book are visible in each page image is
important for preserving potential paratextual material and requires employing a variable
crop width to accommodate the variety in horizontal space between a page’s edge and
the edge of the book’s cover;
(b) Orientation challenges: the lids of flatbed scanners and glass elements of cradles used
to avoid distortion in camera-based digitization equipment can splay pages, making it
difficult to preserve the rectilinear orientation of the text block that contributes to OCR
quality for many engines;
(c) Inconsistent or implausible color: the range of equipment and facility setups used by
partner archives led to uncertainty about color accuracy and some instances of
inconsistent color within a work, leading to inefficiencies in post-processing and
impacting our ability to meet project digitization guidelines;
(d) Lack of confidence in post processing: even when post-processed target images passed
FADGI 3 standards, post-processing proved challenging and time-consuming when
targets were not captured with sufficient frequency or when underlying issues with
equipment impacted certainly regarding color accuracy. Centralizing post-processing at
Emory proved inefficient and introduced additional uncertainties because those
conducting post-processing lacked access to the physical works for comparison.

(6) Ran OCR using GCV;
(7) Created derivatives and access copies of files.

Outcomes

Partner archives digitized twenty-two songbooks, including five to six books from each institution, and
generated high resolution TIFF files with derivative JPEGs for access. The project team also conducted
OCR to obtain text for each work using Google Cloud Vision.

The project team also developed a strategy for digitization, OCR, and MIR for the future implementation
phase. Under this strategy, holdings at partner institutions with robust digitization operations will be
digitized by in-house specialists. Collections at institutions without the capacity to perform efficient mass
digitization in accordance with project standards will be digitized by an industry-leading vendor with a
track record of high-quality work. This hybrid approach to digitization will ensure productive workflows
and high-quality, consistent images for the digital library. Although our approach to OCR will be informed
by the research performed during the planning process, we expect that the landscape will have shifted by
our implementation phase start date. To ensure that we incorporate up-to-date OCR research and tools
into implementation, we will conduct a new study of leading available commercial and open source OCR
engines midway through implementation. Although OMR remains inaccurate and labor-intensive, we will
seek to include more readily obtainable information about the location of music on books’ pages in the
digital library, enhancing the collection’s research potential.
Pilot Site

Context

The siloing of digital collections—frequently accessible only through institutional library catalogs or digital repositories with varied exposure to search engines—is a major factor limiting knowledge or use of collections beyond their institutional homes. Holdings are more discoverable when presented in midsize thematic research collections organized by collection stewards or scholarly collectives. In addition to aggregating digitized works that span institutional holdings, the production of a pilot site for the Sounding Spirit Digital Library offered the project team an opportunity to think through design questions that the development of a thematic collection presents: how best to display works and associated metadata in order to foreground material of greatest interest to the project’s audiences and how to organize the digital library in a way that offers points of access to interested scholars and members of the public. These considerations about what to privilege in organization and content have the potential to shape the accessibility and navigability of the thematic collection. In addressing these concerns, the project was guided by research questions about the role of music books in southern vernacular sacred singing genres and by our interest in serving a diverse audience that includes both members of the public interested in these types of music and scholars interested in using the library’s works for research and teaching. We ultimately chose to foreground a limited selection of metadata that highlights aspects of the works most relevant to these audiences or related to specific features of the corpus, and to organize works into overlapping collections foregrounding place, genre, population, context, or format.

The selection or development of a platform for a thematic collection can also impact access and engagement with the resource. Our selection of the ECDS-developed Readux software as the platform for the digital library enabled us to leverage features of the tool that support research and teaching and to minimize the resources required for site development. Readux offers easy-to-use software for building thematic collections. Other successful thematic collections have typically been one-off custom sites. Such collections may host their own content, as the William Blake Archive or Nineteenth-Century Disability: Cultures and Contexts do, or may collect, organize, and encourage engagement with material stored elsewhere, as the Yiddish Digital Library does with volumes hosted by the Internet Archive. Yet all are built with custom software and required considerable grant funding to develop. Readux lowers the bar as open source software that facilitates teaching and research with thematic collections that span institutional holdings.

Process

(1) Selected ECDS’s Readux platform for its features supporting browsing and engagement, including:
   (a) Deep zoom;
   (b) Transparent overlay of selectable text on top of page images, drawing on text and positional data from OCR;
   (c) Adoption of the International Image Interoperability Framework (IIIF) to support federated collections, interoperability, and addressability;
   (d) Note-taking and export features helpful for teaching;
   (e) Organizational features, including ability to group works into collections, develop static pages, and feature works and collections on a site home page.

(2) Developed organizational schema, drawing on conversations with stakeholders:
(a) Developed plan for metadata display;
(b) Created list of collections to group works on the pilot site;
(c) Created models for descriptive entries, including volume summaries for individual works and collection descriptions that connect works or collections to the project’s key humanities themes of race, place, religion, genre, and modernity.
(3) Drafted descriptive summaries and collection descriptions;
(4) Assembled metadata;
(5) Ingested volumes and metadata into Readux;
(6) Launched site;
(7) Submitted site volumes to the DLG for ingest;
(8) Promoted site via social media and to stakeholders.

Outcomes
The project team launched the Sounding Spirit Digital Library pilot site (https://library.soundingspirit.org) on April 30, 2020, at the conclusion of the yearlong planning process. The pilot site includes twenty-two digitized books organized into twenty-three overlapping collections. Each work in the pilot digital library features fully searchable text, robust metadata, and a volume summary placing the work in thematic and historical context. Each collection also features an overview describing the relationship between vernacular sacred music publishing and the denomination, place, format, musical style, or population represented in the collection.

Future Phases

Context
Building on the success of the yearlong planning process outlined above, the Sounding Spirit project team developed a plan for implementation. Our team’s work conducting bibliographical research, fostering institutional partnerships, creating technical workflows, and developing a pilot digital library provided invaluable experience that informed our design of future phases of the project.

Bibliographic research uncovered a dramatic range of works representing the sacred music-making of diverse populations, leading us to propose digitizing over 1,250 works, rather than the 500 to 700 volumes we initially anticipated including in the implementation phase. Our pilot project also informed our decision to include contextualizing descriptive entries such as volume summaries and collection descriptions in the digital library that increase the accessibility of the resource. Insights gleaned from working with partners representing varied institutional affordances and a wide range of technical skills led us to consider a hybrid approach to digitization in which institutions with robust digitization programs will undertake this work on site while partner institutions with fewer resources and consumer grade equipment collaborate with a third-party vendor to contribute volumes to the digital library. This hybrid approach to digitization will ensure productive workflows and high-quality, consistent images for the larger scale digitization that implementation of the digital library entails. The planning grant process also demonstrated the efficacy of large-scale digitization projects that pair scholarly collectives with archival partners to build thematic collections spanning institutional holdings. Finally, building a pilot digital library equipped us to plan the development and organization of the digital library infrastructure with greater confidence, placing a high value on research, teaching, public engagement, and access.
**Process**

Sounding Spirit’s project team cemented partnerships with continuing and new archival partners and scholarly and technical advisors, and built on a yearlong planning process to develop a list of digitization candidates, a plan for associated descriptive writing, and resources to support the collection of metadata and the further development of the digital library.

(1) Institutional involvement:
   (a) Surveyed partner archives to gauge their capacities and ensure the alignment of the Sounding Spirit project with institutional priorities;
   (b) Recruited two further partner archives based on gaps in the holdings of the original four institutions:
      (i) The University of Tulsa’s McFarlin Library Department of Special Collections and University Archives and Helmerich Center for American Research at the Gilcrease Museum were recruited due to their rich collections of Native American hymnody;
      (ii) The Collections Research Center and William L. Clements Library at the University of Michigan was recruited because of its digitization unit’s experience experimenting with IIIF and its collection of art song arrangements of spirituals and otherwise omitted works;
   (c) Identified candidate volumes for digitization, deriving a list of 1,284 items from the “Checklist of Southern Sacred Music Imprints, 1850–1925” based on collection priorities and the capacities of each institution (see Appendix G);
   (d) Deferred condition assessment of individual volumes to the implementation stage.

(2) Descriptive entries (see Appendix H):
   (a) Developed a collection schema encompassing 100 thematic collections of candidate volumes;
   (b) Dedicated resources to the research and writing of summaries that contextualize individual volumes;
   (c) Planned topics for fifteen bibliographic essays;
   (d) Planned to include pedagogical resources such as primary school lesson plans in the implementation of the project.

(3) Advisors, technical considerations, and budget:
   (a) Recruited additional advisory board members to expand the disciplinary orientations, content expertise, and technical knowledge represented by original board and project personnel;
   (b) Developed a plan with advisory board members to conduct MIR and maximize efficacy of OCR by conducting a new study midway through implementation;
   (c) Obtained quotes from three leading digitization vendors;
   (d) Selected competitive quote from one digitization vendor for use by partner institutions that are unable to perform in-house digitization at scale;
   (e) Submitted a grant application for implementation in July 2020, including a comprehensive budget to accommodate project priorities.

**Outcomes**

ECDS applied for implementation support from the NEH Humanities Collections and Reference Resources program in July 2020. The planned implementation of the digital library will create a virtual humanities
collection providing a single point of access to an additional 1,284 significant books of vernacular sacred music from the southern United States published between 1850 and 1925 and support rich engagement with these seminal, yet under-collected works. Our project plan for implementation includes: 1) Digitization of approximately 1,284 books from six partner archives; 2) standardization of descriptive metadata for all volumes; 3) research and writing of 425 volume summaries and 100 collection descriptions; 4) development of the digital library site, including limited customization of its Readux platform to support domain-specific needs of the project; and 5) pilot activities demonstrating interpretation and impact of this humanities collection, including fifteen bibliographic essays, pedagogical materials, and public programming that encourage teaching and engagement with the library’s contents.

In the longer term, we plan to draw on the works in the Sounding Spirit Digital Library to engage in and support research, education, and public programming connecting southern sacred vernacular songbooks to American history and culture. In conjunction with the launch of the digital library, we expect to convene moderated conversations, hold physical exhibitions of books at partner institutions, and present both concert and participatory performances of music found in the library’s texts. In subsequent phases, the Sounding Spirit initiative plans to curate and present additional events, exhibitions, and media productions.

Future phases of the Sounding Spirit Digital Library may also expand the resource’s coverage to more fully capture the role of vernacular sacred music books in American history and culture by including music books published outside the 1850–1925 range and those associated with additional American and transnational geographies. We are also exploring how we might deploy the thematic collection production strategies we have developed in other topical areas of interest to scholars engaged in research with American music sources, such as those involved in the Society for American Music and the Committee on the Publication of American Music. In these later phases of the initiative, the project team will continue to partner with and mentor others building programming on the foundation laid by the Sounding Spirit Digital Library.

Recommendations in Detail

Our experience planning the Sounding Spirit Digital Library led us to appreciate the value of developing digital thematic collections of humanities materials, forging collaborations among scholars and archives to develop digital resources, and extracting information from digitized humanities materials by applying optical character recognition and related technologies and publishing such information in digital resources. In this section, we articulate why these undertakings are valuable and offer recommendations for how to build such collections, partnerships, and technical workflows.

Build Digital Thematic Research and Teaching Collections

Why?

Thematically organized digital collections of primary sources that support research and teaching have emerged as a growing genre of scholarly publishing. These collections provide an avenue for increasing public engagement with humanities materials and help realize the potential of digitization for enhancing
research with primary sources. Frequently described as “thematic research collections,” such resources can be equally impactful for teaching. Archives should participate in thematic collection development to encourage engagement with their significant holdings. Scholars should participate in the development of thematic collections to further the impact of their work, making the sources underpinning their scholarly interventions more widely accessible to teachers, researchers, and interested publics.

Although digitized works are often nominally accessible through open access institutional repositories, such collections are often little known and underutilized. Even when included in large-scale aggregators such as HathiTrust or the DPLA, such resources can easily be overlooked. Thoughtfully organized thematic collections can realize access potential by attracting audiences who seek out materials related to a topic of interest. Works in such collections can span institutional holdings and be more discoverable thanks to curation and organization. Thematic collections can catalyze research at a scale beyond that which individual scholars can accomplish. Those engaged in developing thematic collections should consider how their design, features, and the inclusion of supplemental materials can help realize their pedagogical and research potential.

How?

- In developing criteria for thematic collections’ scope, start with research questions. Driving the conceptualization of a thematic collection by working from key research questions will help ensure that the planned project makes a significant intervention in humanities scholarship.
  - Develop scope and related criteria by drawing on research questions;
  - Prioritize criteria for inclusion, informed by both research questions and anticipated needs of users;
  - Instrumentalize prioritized criteria by identifying related search terms before beginning to compile catalog records.
- Create pathways into thematic collections by considering their organization, metadata, and descriptive content.
  - Draw on scholarly interpretive framework to foreground organization around themes that will provide curated points of entry into the collection;
  - Identify domain-specific descriptive metadata to facilitate user experience and enhance research and teaching potential;
  - Digital thematic research and teaching collections can lie at the intersection between humanities collections and reference resources; consider including descriptive entries—such as summaries of included works or groupings connected to interpretive frameworks—to motivate interaction by offering engaging context;
  - Facilitate teaching and public engagement by developing pedagogical materials and modeling public programming that draws on the thematic collection.
- Facilitate interoperability and addressability through digitization workflows and metadata. Individual researchers, content aggregators, and additional digital resource initiatives can study and enhance the reach of works in thematic collections or institutional repositories if they are published using protocols that support interoperability and computational reuse.
  - Include OCR with textual and positional information, transcriptions, or other computationally recognized information alongside primary sources;
  - Adopt IIIF or other emerging standards to facilitate inclusion in federated collections, harvesting, and computational reuse;
○ Publish harvestable descriptive metadata that adopts standards published by aggregators (such as state partners of the DPLA) to enhance the discoverability of works within the thematic collection.

- Adopt user-friendly and feature-rich tools to realize potential for research, teaching, and public access. The open source Readux platform adopted by Sounding Spirit offers an economical model for encouraging engagement through its inclusion of annotation and publishing features.

- Share your work.
  ○ Publish the bibliographic research conducted to facilitate the development of a thematic collection; such research includes information of value to researchers, whether or not works were selected for digitization, and can support future bibliographic undertakings;
  ○ Include a data dictionary to ensure that others can correctly interpret bibliographic research;
  ○ Publish process documents that might help others engaged in related projects, including documents that helped guide the selection of works included in the thematic collection and those used to survey and record workflows among partners.

Foster Collaboration in Digital Collection Development

Why?
Thematic digital research and teaching collections will be most successful when they emerge from collaborations—convening archives with complementary collection strengths, scholars with related research interests or topical foci, and digital scholars and technologists with expertise in digitization workflows or the design and technical considerations of interfaces for engaging with digitized texts. Archives that form partnerships with public scholars and other institutions will create digital collections that more successfully meet the interests of scholarly and public audiences. Scholars who work collaboratively will be able to bring complementary subject matter expertise to bear on collections’ scope, organization, and descriptive content. Teams drawing from these diverse domains will be well-positioned to seek funding or to pool resources, making the development of thematic collections more broadly achievable.

How?
- For archives and collection stewards:
  ○ Identify institutions with holdings that complement collection strengths or priorities for digitization, and seek partnerships;
  ○ Develop shared criteria and priorities for collaborative thematic collections;
  ○ Develop shared bibliographic research methods and data standards to survey holdings.

- For scholarly collectives:
  ○ Work from topics of mutual interest or diverse research questions connected to a common body of primary sources to identify potential scopes for thematic collections;
  ○ Identify institutions with key holdings and relevant expertise.

- For digital humanities centers:
  ○ Develop reusable platforms, lowering the technical and resource requirements to develop thematic collections;
Partner with scholars and archives to identify initiatives to guide the development of such technologies.

Accommodate Institutions with Diverse Affordances in Digitization and OCR Workflows

Why?

Digitization initiatives are among the projects “that provide an essential underpinning for scholarship, education, and public programming in the humanities,” as articulated in the NEH Humanities Collections and Reference Resources program description. Digitization protocols that weigh both standards and project use cases facilitate the participation of institutions with diverse affordances. Including robust OCR in digitization workflows can make digital surrogates more accessible and useful to varied audiences including through computational research.

How?

- Adopt federal or international digitization guidelines, but consider deviations when justified by intended use cases, if these enable the participation of lower-resourced institutions. Guidelines such as those of FADGI, are well-conceived and valuable, but may be expensive to achieve and beyond the capabilities of many institutions. Consistency and quality are worthy goals, but so is inclusion.
  - Consider intended outcomes and audiences when adopting digitization standards;
  - Consider allowing for flexibility within defined ranges around some specifications if this will enable the participation of lower-resourced institutions with significant holdings that might not otherwise be included in thematic collections.
- Partner with a vendor when doing so supports cost-effective collaboration. Collaborative collections—or those anchored by institutions without the capacity to carry out FADGI-compliant digitization onsite—can achieve consistent and high-quality results through engaging digitization vendors.
  - For institutions with limited staffing, consumer-grade equipment, or facilities that pose challenges for digitization, consider seeking or allocating funds to employ vendors who can offer cost-effective digitization services.
- Institutions with robust digitization capacity can and should continue to build capacity and perform high-quality digitization on site. Such institutions can frequently conduct digitization at comparable quality and lower costs than leading vendors.
  - Institutions should maintain and continue to develop expertise, equipment, and facilities that support such work;
  - Institutions should consider offering services that might support the development of digital collections that include works from the holdings of lower-resourced institutions, such as supporting digitization of works held by other institutions as part of prioritized or funded projects, or via a fee-for-service model.
- Digitization programs should collect and publish positional information in addition to textual content. Such information can be harnessed for research or can be deployed by platforms for browsing digitized works, like Readux, to enrich user interaction.
• Archives and other collaboratives should regularly reassess OCR and related practices to take advantage of rapid advances in OCR accuracy.
  o Archives working with digitized collections that include particular features (such as printed music) should consider adopting MIR or other domain-specific computer vision technologies to obtain and publish useful information;
  o Archives should consider building staff skill sets to take advantage of leading cloud-based commercial computer vision services or open source OCR engines. Commercial engines produce accurate OCR at low cost. Open source engines frequently require training on reference corpuses but are free to use and can generate accurate results. Both sets of tools require specialized skills to employ effectively, but are considerably more affordable and accurate than legacy software still in use at many archives;
  o Digital scholarship centers should develop open source software that makes applying cutting-edge OCR engines more user friendly.
Appendix A: Prioritized Criteria and Search Processes

Prioritized Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Range</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of Publication</td>
<td>1850–1925.</td>
<td>May include major editions of works published before 1850 that remained in print and active circulation after 1850.</td>
</tr>
<tr>
<td>Geographic Origin</td>
<td>US South.</td>
<td>May include locales connected to the South via migration. May include places of compilation/composition and printing/publication.</td>
</tr>
<tr>
<td>Sacred Content</td>
<td>Any sacred music books or books within geographic and temporal scope.</td>
<td>Will include many Protestant books but also any Catholic, Jewish, Muslim, or other sacred publishing.</td>
</tr>
<tr>
<td>Vernacular Content</td>
<td>Exclude most art music.</td>
<td>Not typically reflected in catalog records.</td>
</tr>
<tr>
<td>Partner Archives</td>
<td>Emory University, Middle Tennessee State University, Southern Baptist Theological Seminary, University of Kentucky.</td>
<td>Leverage strengths of partners by exhaustively surveying their relevant holdings, but also gauge what the key gaps are and conduct bibliographic research to identify significant or underrepresented works that aren’t held by partners. We can identify additional partners for implementation.</td>
</tr>
<tr>
<td>Race and Ethnicity</td>
<td>All works regardless of race or ethnicity.</td>
<td>Seek to include publications representing Native American and Black compilers, contributors, and communities of use, as well as those associated with immigrant or other non-white populations.</td>
</tr>
<tr>
<td>Genre</td>
<td>All books regardless of genre.</td>
<td>Relevant genres include: Collections of spirituals, Sunday school songbooks, words only hymnals, denominational hymnals, shape-note tunebooks, and gospel songbooks.</td>
</tr>
<tr>
<td>Language</td>
<td>All books regardless of language.</td>
<td>Seek to include non-English language books, including those in Native languages, those in Spanish, German, French, or any other languages associated with immigrant populations.</td>
</tr>
</tbody>
</table>
Influence and Significance
Influential texts in the area published out of region.
Include the most important texts published out of region that were used in the South. Make a special effort to identify books that are unique or local productions.

Availability
Books that are not presently available in high-quality open access digital editions.
Books that are not presently available digitally or in print or that are not widely collected in libraries are modestly preferable to those that are more widely accessible.

Search Process

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Process</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of Publication</td>
<td>Search for publication dates. (Undated material will still return on a subject-based search.)</td>
<td>1850 to 1925.</td>
</tr>
<tr>
<td>Geographic Origin</td>
<td>Keyword search for town names, especially county seats (older records will also include abbreviated state names).</td>
<td>Locations of interest: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, Washington (DC). Potential locations of interest: Maryland, Missouri, Ohio Valley area. Locations of interest related to the Great Migration: Baltimore, Boston, Buffalo, Chicago, Cincinnati, Cleveland, Columbus, Detroit, Indianapolis, Kansas City (MO), Los Angeles, Milwaukee, New York City, Oakland, Philadelphia, Phoenix, Pittsburgh, St. Louis, San Francisco, Seattle, Toledo, Washington (DC). Locations of interest related to Indian Removal: Indian Territory.</td>
</tr>
<tr>
<td>Sacred Content</td>
<td>Search subject headings for Christian denominations.</td>
<td>Subject headings for Christian denominations: Adventists—Hymns; African American Baptists—Hymns (also &quot;African American Baptists—[state name]—Hymns&quot;); African American churches—Hymns; African Methodist Episcopal Church—Hymns; Amish—Hymns; Anabaptists—Hymns; Anglican Communion—Hymns; Assemblies of God, General Council—Hymns; Associate Reformed Presbyterian Church—Hymns; Baptist hymns; Baptists—Hymns (and &quot;Baptists—[state name]—Hymns&quot;); Bohemian Brethren—Hymns; Brethren in Christ Church—Hymns; Bruderhof Communities—Hymns (probably arrived in US too late to be of use); Catholic Church—Hymns (and &quot;Catholic Church—[state/region/rite]—Hymns&quot;);</td>
</tr>
</tbody>
</table>
Catholic Hymns; Christian and Missionary Alliance—Hymns; Christian Reformed Church—Hymns; Christian Science—Hymns; Church of God—Hymns (with variations by conference); Church of God, Holiness—Hymns; Church of Jesus Christ (Bickertonites)—Hymns; Church of Jesus Christ of Latter-Day Saints—Hymns; Church of the Brethren—Hymns; Church of the Nazarene—Hymns; Church of the United Brethren in Christ—Hymns (variations by conference); Churches of Christ—Hymns; Churches of God in North America—Hymns; Community of Christ—Hymns; Congregational churches—Hymns; Congregationalists—Hymns; Cumberland Presbyterian Church—Hymns; Disciples of Christ—Hymns; Divine Science Church—Hymns; Dunkard Brethren—Hymns; Episcopal Church—Hymns; Evangelical and Reformed Church—Hymns; Evangelical Association of North America—Hymns; Evangelical Congregational Church—Hymns; Evangelical Covenant Church of America—Hymns; Evangelical Lutheran Church in America—Hymns (not certain when they and the Missouri Synod split—may be too late); Evangelical Protestant Church of North America—Hymns; Evangelical Reformed Church—Hymns; Evangelical Synod of North America—Hymns; Evangelical United Brethren Church—Hymns; Free Baptists—Hymns; General Church of the New Jerusalem—Hymns; General Conference Mennonite Church—Hymns; General Convention of the Christian Church—Hymns; General Synod of the Evangelical Lutheran Church in the United States—Hymns; Greek Orthodox Archdiocese of North and South America—Hymns; Heavenly Recruit Association—Hymns; Holiness churches—Hymns (and "Holiness churches—[state]—hymns"); Huguenots—Hymns; Hussites—Hymns; Hutterite Brethren—Hymns (and "Hutterite Brethren—[state/region]—Hymns"); Hymns—Baptists; Hymns, Lutheran; International Association of Religious Science Churches—Hymns; International Churches of Christ—Hymns; International Society of Christian Endeavor—Hymns; Israel of God’s Church White Horse Army—Hymns; Jehovah’s Witnesses—Hymns; Jesuits—[state]—Hymns; Liberal Catholic Church—Hymns; Lutheran Church—Missouri Synod—Hymns; Lutheran Church—United States—Hymns; Lutherans—Hymns; Mennonite Brethren in Christ—Hymns; Menlnonite Church—Hymns; Mennonites—Hymns (and "Mennonites—[state/region]—Hymns"); Methodism—Hymns; Methodist Church—Hymns; Methodist Episcopal Church—Hymns; Methodist Episcopal Church, South—Hymns; Methodists—Hymns; Moravian Church—Hymns;
<table>
<thead>
<tr>
<th>Vernacular Content</th>
<th>Exclude most art music identified through</th>
<th>Publishers of art music whose works should typically be excluded: Ditson, Schirmer, co-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moravian Church in America—Hymns; Mormon Church—Hymns; Mormons—Hymns; National Baptist Convention of the United States of America—Hymns; National Primitive Baptist Convention of the U.S.A.—Hymns; New Jerusalem Church—Hymns; Old German Baptist Brethren—Hymns; Old Order Mennonite Church—Hymns; Old Regular Baptists—Hymns (and &quot;Old Regular Baptists—[state/region]—Hymns&quot;); Orthodox Eastern Church—Hymns (and &quot;Orthodox Eastern Church—United States—Hymns&quot;); Pentecostal Holiness Church—Hymns; Pentecostals—Hymns; Presbyterian Church—Hymns; Presbyterian Church in the U.S.A.—Hymns; Presbyterian Church (U.S.A.)—Hymns; Presbyterian Hymns; Presbyterians—Hymns; Primitive Baptists—Hymns; Primitive Methodist Church—Hymns (mostly going to be British); Protestant churches—Hymns (and &quot;Protestant churches—[state/region]—Hymns&quot;); Quakers—Hymns; Reformed Church—Hymns; Reformed Church in America—Hymns; Reformed Church in the United States—Hymns; Reformed Episcopal Church—Hymns; Reformed Presbyterian Church of North America—Hymns; Reformed Protestant Dutch Church (U.S.)—Hymns; Reorganized Church of Jesus Christ of Latter Day Saints—Hymns; Restoration movement (Christianity)—Hymns; Salvation Army—Hymns; Schwenkfelder Church—Hymns; Schwenckfelders—Hymns; Seventh-Day Adventists—Hymns; Seventh-Day Baptists—Hymns; Seventh-Day German Baptist Brethren—Hymns; Shakers—Hymns; Society of Christian Israelites—Hymns; Society of Friends—Hymns; Society of Separatists of Zoa—Hymns; Southern Baptist Convention—Hymns; Theosophy—Hymns; Unitarian Church—Hymns; Unitarian Universalist Association—Hymns; United Baptists—[state/region]—Hymns; United Brethren in Christ—Hymns; United Church of Christ—Hymns; United Church of Religious Science—Hymns; United Lutheran Church in America—Hymns; United Methodist Church (U.S.)—Hymns (I think they merged in the 1950s); United Presbyterian Church in the U.S.A.—Hymns; United Presbyterian Church of North America—Hymns; Unity School of Christianity—Hymns; Universalist Church—Hymns; Worldwide Church of God—Hymns; Zion Christian Church—Hymns</td>
<td>Subject headings for non-Christian hymns: English [religion name] hymns; Hymns, Islamic; Hymns, Muslim; Jewish hymns, Synagogue music.</td>
<td></td>
</tr>
</tbody>
</table>
| **Partner Archives** | For individual archives:  
1) see if a checklist or shelf list of hymnody-specific collections is available or can be returned;  
2) failing that, first conduct a date-delimited search of the "Hymns, English" subject heading (and subsequently other languages of interest);  
3) read through records by hand to check places of publication or distribution and note down those of interest;  
4) repeat process to locate records not cataloged as "Hymns" using other subject headings associated with criteria such as "Spirituals (songs)," "Gospel music," etc.;  
5) make special arrangements to consult only locally cataloged material (applicable to MTSU and SBTS).  
For record aggregators like OCLC:  
1) conduct searches by locations, denominations, and other categories detailed to identify works not held by partner institutions;  
2) note down sigla of institutions that have marked holdings of works of interest. |
| **Race and Ethnicity** | Search subject headings for Native languages with Native languages with authorized headings for hymns: Hymns, Cherokee; Hymns, Cheyenne; Hymns, Chilliwack; Hymns, Chinook jargon; Hymns, Chipewyan; Hymns, Chippewa; Hymns, Choctaw; |
| **headings for hymns** (regardless of region). Searches subject headings for black denominations (listed under search criteria for “Sacred”). Searches subject headings for denominations and languages with headings for hymns associated with immigrant groups. | Hymns, Cree; Hymns, Creek; Hymns, Dakota; Hymns, Delaware; Hymns, Eskimo; Hymns, Hawaiian; Hymns, Inuit; Hymns, Kiowa; Hymns, Kipchak; Hymns, Mixtec; Hymns, Mohawk; Hymns, Munsee; Hymns, Nahuatl; Hymns, Navajo; Hymns, Nez Perce; Hymns, Ojibwa; Hymns, Oneida; Hymns, Passamaquoddy; Hymns, Pawnee; Hymns, Santee; Hymns, Seneca; Hymns, Shuswap; Hymns, Sioux; Hymns, Spokane. |
| **Genre** | Subject headings associated with relevant genres: Spirituals (songs); Sunday school music; Hymns—juvenile; Hymns, [language]—juvenile; Shape-note hymnals; Gospel music; Revival hymns. |
| **Language** | See list of Native languages under “Race and Ethnicity.” Subject headings associated with immigrant populations: Hymns, Czech; Hymns, Dutch; Hymns, French; Hymns, German; Hymns, Hebrew; Hymns, Portuguese; Hymns, Spanish. |
| **Influence and Significance** | Work with advisory board members and other subject matter experts to identify out-of-region works that merit inclusion or to identify local publishers to track. |
| **Availability** | When searching individual catalogs, search titles of interest to see whether there are high-quality open access versions available. When searching OCLC, check for presence “Downloadable Musical Score” or “eBook” formats for collections of interest. |
Appendix B: Metadata Schema

Metadata Schema

This document outlines the draft descriptive metadata schema for the Sounding Spirit digital library. The table below provides a description of each field, offers information about its values and their source, notes whether a field corresponds to a Dublin Core or Digital Library of Georgia/Digital Public Library of America field, and indicates whether a field's values rely on a local or external controlled vocabulary. Each work in the Sounding Spirit digital library will also feature technical and preservation metadata generated through an automated process as part of the post-digitization workflow, and administrative metadata tracking elements of this workflow.

Key

* = Corresponds to a Digital Library of Georgia required field.
** = Corresponds to a Digital Library of Georgia optional field.
(DC) = Corresponds to a Dublin Core field.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Values</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>URL*</td>
<td>Volume landing page in Sounding Spirit digital library.</td>
<td>A web address.</td>
<td>Automated</td>
</tr>
<tr>
<td>Short Title</td>
<td>Short title of the work.</td>
<td>For a work with a subtitle, the pre-colon title. For a long title with no subtitle, an identifying subset of the title.</td>
<td>Derived from the Title.</td>
</tr>
<tr>
<td>Title* (DC)</td>
<td>Full title of the work.</td>
<td>The full title, in sentence case, without leading articles.</td>
<td>Bibliographic record for the work.</td>
</tr>
<tr>
<td>Responsibility</td>
<td>Authors or editors of the given work. All known authors or editors are listed in the order found in the catalog record.</td>
<td>One or more names, separated by commas, with first names followed by last. Names are recorded in the same manner (with initials or abbreviations preserved) as they are found. Titles of authors or editors, such as “Rev.” or “Dr.,” are omitted except in the case of married women who had been credited by their husbands’</td>
<td>Bibliographic record for the work.</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
<td>Controlled, external</td>
<td>Library of Congress Name Authority File</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>---------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Creator* (DC)</td>
<td>Authors or editors of the given work. All known authors or editors are listed in the order found in the catalog record.</td>
<td>Controlled, external</td>
<td>Library of Congress Name Authority File</td>
</tr>
<tr>
<td>Place of Publication (DC)</td>
<td>Places associated with the given work, including places of publication, compilation, distribution, or collection of ethnographic material. All known places of publication or distribution are included. When known, the place of compilation and/or collection is listed first, while other places are listed in the order found in the catalog record.</td>
<td>For locations within the United States, cities (when known) are followed by a comma and then by the state's postal code. For locations in Canada, cities (when known), are followed by a comma and then by the province. For other locations outside the United States, cities (when known) are listed alone, with the country name added when it is necessary for purposes of clarity.</td>
<td>Most information derives from the bibliographic record for the work. This information is supplemented with research regarding the work at hand when relevant to the inclusion of the work in the checklist (for instance, regarding material published outside the US South but representing the work of authors based or trained there).</td>
</tr>
<tr>
<td>Coverage (Spatial)*</td>
<td>The name of the place in which the physical item originated, including county and city, if known.</td>
<td>Controlled, external</td>
<td>Getty Thesaurus of Geographic Names</td>
</tr>
<tr>
<td>Date of Publication* (DC)</td>
<td>The date of publication or printing for the work, whichever is later.</td>
<td>A year. Works published serially feature two dates representing the beginning and ending of the period of publication separated by a forward slash. When the work was issued without a date, approximate dates supplied by the cataloger appear with the uncertain digits replaced by question marks, as in “191?” or “18??.” Dates preceded</td>
<td>Getty Thesaurus of Geographic Names</td>
</tr>
<tr>
<td><strong>by the term “bef.” indicate that for a work issued without a date, the cataloger has estimated that the work was published before the year indicated.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Publisher</strong> (DC)</td>
<td>The name of the publisher or publishers that issued the work.</td>
<td>The name of the publisher or publishers as it appears in the record consulted, with formatting and abbreviations preserved. For works with multiple publishers listed, each entry is separated by semicolons.</td>
<td>Bibliographic record for the work.</td>
</tr>
<tr>
<td><strong>Language</strong> (DC)</td>
<td>The language or languages in which the work is written.</td>
<td>The name of the language or languages. If multiple languages appear in the same work, each language name is separated by semicolons. If English and another language appear in the same work, the non-English language has been listed first for ease of sorting. The designation is followed by a question mark if the determination about the work’s language is uncertain.</td>
<td>Bibliographic record for the work, supplemented by examination of the original when needed.</td>
</tr>
<tr>
<td><strong>Language Abbreviation</strong></td>
<td>Three-digit code(s) of the language or languages in which the work is written.</td>
<td><em>Controlled, external</em></td>
<td>ISO 639-2</td>
</tr>
<tr>
<td><strong>Pagination</strong> (DC)</td>
<td>An indication of the numbering sequence of pages in the work</td>
<td>For paginated works, a number indicating the overall number of pages in the work, with any sequences of separate pagination indicated by the insertion of a comma; for unpaged works, a number indicating the number of pieces appearing in the work.</td>
<td>Bibliographic record for the work, supplemented by examination of the original when needed.</td>
</tr>
<tr>
<td>Size (DC)</td>
<td>The height and width of the physical work.</td>
<td>The height of the work as measured from the top edge or head to the bottom edge or tail of the cover and the width of the work as measured from the spine edge to the fore edge of the cover, in centimeters, separated by an “x.”</td>
<td>Bibliographic record for the work, supplemented by original measurements when needed.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Extent**</td>
<td>The number of digital images in the digitized version of the work.</td>
<td>An integer.</td>
<td>Automated</td>
</tr>
<tr>
<td>Summary* (DC)²</td>
<td>A description of the work, its makers, and its context.</td>
<td>A 150-word summary.</td>
<td>Researched and written by a subject matter expert recruited by the project team.</td>
</tr>
<tr>
<td>Notation</td>
<td>The type of music notation, if any, used in the work.</td>
<td>Controlled, local Works with no notated music are indicated with “Nil.” For works that do contain notated music, “Round” indicates the use of round or standard notation, “Seven-shape” and “Four-shape” the use of shape-notes, and “Numeral” and “Numeral staff” the use of forms of numeral notation. “Mixed” indicates the use of multiple systems of notation.</td>
<td>Bibliographic record for the work, supplemented by examination of the original when needed.</td>
</tr>
<tr>
<td>Content</td>
<td>An indication of whether the work contains secular material in addition to sacred material</td>
<td>Controlled, local “Sacred” if the work’s contents are sacred in nature; “Mixed” if the work is also known to contain secular material.</td>
<td>Bibliographic record for the work, supplemented by examination of the original when needed.</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Designations of whether the work meets certain criteria of interest to scholarly</td>
<td>Controlled, local Values are: Broadside; Convention Gospel; Confederate Imprint;</td>
<td>Bibliographic record for the work, supplemented by examination of the original when needed.</td>
</tr>
</tbody>
</table>

² Corresponds to the Dublin Core and DLG fields titled “Description.”
and public audiences pertaining to format, genre, identity, geography, use, and other themes.

<table>
<thead>
<tr>
<th>Subject* (DC)</th>
<th>Subjects covered by the work.</th>
<th>Controlled, external Separated by semicolons.</th>
<th>Library of Congress Subject Headings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributing Archive**3</td>
<td>The name of the archive holding the physical copy of the work.</td>
<td>Controlled, local</td>
<td>Bibliographic record for the work.</td>
</tr>
<tr>
<td>Holding Institution*</td>
<td>The name of the institution responsible for the digitized version of the work.</td>
<td>Controlled, local</td>
<td>Automated</td>
</tr>
</tbody>
</table>

3 Corresponds to the DLG field “Original Collection.”
Appendix C: Digitization Processes Questionnaire

(1) Review the Digital Library Use Case carefully (see Appendix E). Where will your institution be able to comply with the project’s digitization standards? Where will your institution need support?

(2) How does your institution digitize books? Please include links and/or copy of all documented processes and protocols pertaining to digitization workflow.

(3) What equipment does your institution own for digitizing books? Please provide product names and numbers and include specifications relevant to our use case. Include items such as:
   (a) Scanners
   (b) Cameras (both camera body as well as lenses used)
   (c) Capture software (e.g. Lasersoft Silverfast, VueScan, Capture One Heritage Edition, Epson Scan, etc.)
   (d) Operating system(s) in use (Windows, Mac, Linux, etc.)
   (e) Color targets

(4) How does your institution process digital files post-digitization? What file types does your digitization process generate? What quality standards do you adhere to?

(5) What capacities does your institution have to run optical character recognition (OCR) on digitized books? What kinds of outputs is your OCR software capable of producing (e.g. searchable PDF, plain text files, ALTO, etc.)?

(6) What partnerships/relationships does your institution support to extend its digitization capabilities (outsourcing, etc.)?

(7) Who at your institution undertakes digitization, OCR, and other post-processing? How many people are involved? How is this effort staffed? How would your institution accommodate a scaled-up digitization effort?

(8) What does your institution do to share and preserve digitized books? Does your institution have a digital repository? If so, what platform is it built in? Does your institution share digitized files with the Internet Archive or HathiTrust? Does your institution implement IIIF or plan to?

(9) What research on digitization, OCR, or digital preservation and access would you recommend other project team members read?
Appendix D: Memorandum of Understanding Template

Memorandum of Understanding

Between Emory Center for Digital Scholarship and [PARTNER LONG].

I. Purpose

This Memorandum of Understanding ("MOU") outlines the relationship between Emory University, through the Emory Center for Digital Scholarship ("ECDS"), and [PARTNER INSTITUTION] ("[PARTNER]") for the Sounding Spirit digital library planning project ("Sounding Spirit Planning") for future implementation ("Sounding Spirit Implementation").

The parties agree that while the matters covered in this agreement are meant to address present areas of mutual concern, it is understood that this memorandum is not all-inclusive, but serves as a statement of intent and not an enforceable contract. Other matters may arise that require mutual discussion and agreement, and other areas of collaboration may be identified. These shall be addressed on a case-by-case basis.

II. Description of Sounding Spirit Digital Library Planning Project

Sounding Spirit Planning is a collaborative year-long planning process beginning on May 1, 2019, and concluding on April 30, 2020, for a project building thematic collections of digitized songbooks that showcase how communities used texts to navigate race, place, religion, and culture across American geographies. Sounding Spirit Planning is a collaboration among ECDS and four contributing archives, including [PARTNER]. This planning process will:

- Build a pilot site featuring a sample selection of southern vernacular sacred songbooks published between 1850 and 1925,
- Identify a list of volumes of southern vernacular sacred songbooks published between 1850 and 1925 to digitize and include in an expanded Sounding Spirit Implementation site if further funding is secured after the conclusion of Sounding Spirit Planning, and
- Conduct an extensive, collaborative research process to identify digitization, optical character recognition (OCR), access, and preservation practices that facilitate inter-institutional collection development.

Sounding Spirit Planning’s digital library pilot site uses the Readux platform, facilitating public access, research, and publishing. The site will launch with twenty volumes exemplifying the strengths of each archive’s collection. Our findings on collaborative digital collection development will be published in a freely accessible white paper and presented at scholarly meetings with the goal of supporting other consortia of scholars and libraries engaged in related work. Sounding Spirit Implementation is dependent on securing additional funding and would follow and build on Sounding Spirit Planning by digitizing...
hundreds of sacred songbooks within the project parameters identified during the planning process and building a Readux site featuring collections of these works.

III. Mutual Commitments

A. ECDS and [PARTNER] will draft, review, and approve a list of volumes housed in the [PARTNER] collections to be digitized for the Sounding Spirit Planning pilot library.

B. ECDS and [PARTNER] agree to draft, review, and approve a list of volumes housed in the [PARTNER] collections that would be digitized and included in Sounding Spirit’s digital library collections if implementation funding is secured.

C. ECDS and [PARTNER] will employ processes and standards for optical character recognition, digitization, and digital preservation reviewed and vetted during Sounding Spirit Planning.

D. ECDS and [PARTNER] will assess and determine funding needs for implementation. Both parties will identify appropriate funding sources to support implementation needs.

IV. Primary Responsibilities

A. Emory Center for Digital Scholarship

1. ECDS agrees to develop and communicate standards for digitization that meet the requirements of Sounding Spirit Planning.

2. ECDS agrees to develop standards for optical character recognition that meet the requirements of Sounding Spirit Planning and to conduct optical character recognition according to these standards on the books digitized by [PARTNER].

3. ECDS agrees to develop and host the Readux site that will feature the volumes digitized by [PARTNER].

4. ECDS agrees to ingest [PARTNER] digitized books into Readux.

5. If the pilot project is successful, ECDS will take lead on applying for funding for Sounding Spirit Implementation.

B. [PARTNER LONG]

1. [PARTNER] agrees to digitize the books in the [PARTNER] collection selected for Sounding Spirit Planning meeting the standards and timeline developed and communicated by ECDS. [PARTNER] agrees to create or supplement basic metadata for these books.

2. [PARTNER] agrees to preserve the files created by digitizing the books in the [PARTNER] collection selected for Sounding Spirit Planning in an institutional repository or other cloud or physical secure backup storage according to the policies of the institution where [PARTNER] is located.

3. [PARTNER] agrees to serve digital files via a IIIF server with positional text compatible with Readux for a period of not less than ten years; OR [PARTNER] agrees to publish digital files via Internet Archive in a manner compatible with Readux for a period of not less than ten years; OR [PARTNER] agrees to share digital files with ECDS and permit ECDS to serve files via a IIIF image server for a period of not less than ten years.

A. [PARTNER] agrees to obtain approval for the digitization of materials at the institution selected for Sounding Spirit Implementation.
V. Term and Termination

This MOU states the parties’ intentions to pursue the project described above, but it is not intended to serve as an enforceable contract. Each party will exert commercially reasonable efforts to fulfill the requirements of this MOU, but each party also acknowledges that changes in funding, staffing, or other unforeseen factors may make it impossible for a party to fulfill all the obligations above. Either party may terminate this MOU if any material breach remains unaddressed for more than thirty days following written notice of the breach from the non-breaching party. Either party may terminate this MOU with or without cause after providing ninety days advance notice.

VI. Authorization

Agreed:

______________________________________
Wayne H. Morse, Co-Director
Emory Center for Digital Scholarship

Date

______________________________________
[REPRESENTATIVE], [ROLE]
[PARTNER LONG]

Date
Appendix E: Digital Library Use Case

Overview

Digitized music books will be published in a digital library in which users may browse them, annotate their words and page regions, and export them as standalone digital editions. Users will include scholars and students wishing to understand the texts’ contents, their production, or their social context. Users will also include participants in music cultures connected to these works who may want to browse the books or bring copies with them to sing. The music books will be displayed in Readux, which enables deep zoom and text highlighting by transparently layering OCR results on top of high-resolution images.

Image Quality

Digital files should offer an accurate and reliable representation of the physical object. Books should be legible and clear when zoomed in.

- Lossless TIFF format
- 400 dpi+ resolution
  - Check for optical resolution spec on scanning device (smaller number than typically advertised)
- Full color
  - 24-bit RGB color depth
  - FADGI 3-star recommends color space Adobe 1998/ProPhoto/eciRGBv2
  - Metamorfoze Light recommends color space Adobe 1998/eciRGBv2
- Sharp focus
  - FADGI 3-star recommends maximum MTF (modulation transfer function) of <1.1
- No visible scratches
- None or minimal warping, bowing, or other distortion
  - Metamorfoze defines geometric distortion as a change in the ratio between vertical and horizontal lines by no more than 2%
  - Best practices:
    - Use of high-quality lenses for image capture
    - Use of distortion-correction software where necessary
- Accurate color
  - Color accuracy should be computed using a color checker device and Digital Imaging Conformance Environment (DICE) software. OpenDICE is an open source version of this software.
  - FADGI 3-star recommends <5 color units difference using ISO 13658:2000
  - Metamorphoze Light also recommends <5 color units difference
  - Best practices:
    - Room should have neutral gray walls
    - Overhead lighting should be 5000K
    - Scanning device should be positioned away from monitors and windows
    - Monitor’s color temperature should match viewing environment
• Little or no distracting information in the shots aside from the images themselves
  o Images should be cropped so that all edges are visible with ¼ to ½ inch of padding
  o Images should be captured or rotated so that the book block is aligned with the image borders
• Content should not be obscured by gutter or digitization apparatus

Federated Collections

Books should be able to be contributed to the digital library from a range of archives with diverse digitization affordances. The digitization process should be scalable to accommodate hundreds of books.

• Be able to accommodate at least some flatbed and camera-based digitization setups
• Be okay with white, black, gray, or other non-distracting backgrounds
• Replicable and relatively efficient capture and post-digitization processing methods

Selectable Text

Text obtained through OCR should have a high degree of accuracy to render it useful but should require a level of effort that is sustainable for a digital library of hundreds of volumes.

• Consistent, high-quality approach to OCR
• ALTO XML or other standard format including positional and textual information
Appendix F: Digitization Workflow

Scope
This document charts the prescribed best practices for the digitization and OCR production for the Sounding Spirit initiative. This process pertains primarily to texts digitized at Emory University, but will address relevant details for partner archives contributing to the pilot Digital Library.

Roles

- Project Director (PD)
- Project Manager (PM)
- Digital Process Supervisor (DPS)
- Digital Production Associate (DPA)
- Music Bibliography Associate (MBA)
- Technical Leads (TLs)

Tasks

<table>
<thead>
<tr>
<th>#</th>
<th>Task</th>
<th>Person</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Confirm specific work to digitize with PD and MBA.</td>
<td>TLs/DPA</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Create card on Sounding Spirit Digitization Trello board.</td>
<td>DPA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Track digitization on Trello.</td>
<td>PM</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Digitization teams submit sample scans to Box for review</td>
<td>TLs/DPA</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>If from ATIZ setup:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Images first staged on Box temp file for post-processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- DPS or DPA develops profile and applies to all pages</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Final image files added to Box for review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>QA 1: Sample Scan Review (see rubric, below)</td>
<td>DPS</td>
<td>300dpi minimum for FADGI 3, 400dpi for FADGI 4</td>
</tr>
<tr>
<td></td>
<td>- Lossless TIFF format</td>
<td></td>
<td><strong>Emory Atiz process:</strong> Begin with roughly 9 Aperture, ½ shutter</td>
</tr>
<tr>
<td></td>
<td>- 600/400 dpi resolution</td>
<td></td>
<td>speed, 200 ISO and modify per book from there.</td>
</tr>
<tr>
<td></td>
<td>- Full color (24-bit)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Accurate color</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Sharp focus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- No visible scratches</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- None or minimal warping/bowing/distortion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Little or no distracting information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Content should not be obscured by gutter or</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
</tbody>
</table>
|6 | Digitization team scans volume using the following best practices:  
  - Clean scanner bed or glass platen  
  - Calibrate color using color checking device and/or scanning software.  
  - Ensure scanning bed is free from distracting information.  
  - Ensure page image is not rotated improperly.  
  - Ensure there are no removable folds or creases in the page.  
  - Strive to capture the full margin of the page in the scan taking care not to damage the book.  
  - Scan all pages at 400 dpi+ in full color.  
  - Save files as uncompressed TIFFs labeled with volume name abbreviation (from digitization tracker) and page number padded with zeroes. (e.g. 1stSix-YYYY-a-0001.tif) **Do not use spaces or underscores in filenames.** | PM |
|7 | Convey review results and request any necessary adjustments, including Emory's completion of all scanning. | TLs/DPA | **Emory post-process:**  
  7a. Confirm all pages have been scanned and ensure correct numbering before leaving digitization station. Save files initially as 1stSix-YYYY-a-0001-Recto/Verso.tif.  
  7b. Save Canon Raw files to Box for review and post-processing.  
  7c. DPS or DPA will process files from CR2 to TIFF by developing unique color profile off color target (see other notes for details). |
|8 | Deliver the uncompressed TIFFs to SoSp via Box. | TLs/DPA |
|9 | QA 2: Full volume review  
  - Ensure the files have the appropriate naming scheme:  
    - First 6 characters of title (excluding “the”)  
    - Year of publication  
    - 4-digit sequence number (e.g. 0001)  
  - Review each scan to make sure that all pages of the original book are present, acknowledging that some pages might be missing per institutional notes.  
  - Ensure images of color targets are present.  
  - Check all volume images for adherence to the following quality requirements laid out in the use case:  
    - Lossless TIFF format  
    - 400 dpi+ resolution  
    - Full color (24-bit where available)  
    - Accurate color  
    - Sharp focus  
    - No visible scratches  
    - None or minimal warping/bowing/distortion  
    - Little or no distracting information | DPS | **Emory post-process:**  
  9a. DPS or DPA will run GoldenThread tests to determine level of FADGI compliance for target images and adjust profile as necessary.  
  9b. Profile is then applied to scans and FADGI compliance determined for select images as required by best practices.  
  9c. We will record details of FADGI pass/fails for later documentation.
<table>
<thead>
<tr>
<th></th>
<th>Content should not be obscured by gutter or digitization apparatus</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Convey review outcomes and request any adjustments to TLs.</td>
<td>PM</td>
</tr>
<tr>
<td>11</td>
<td>Once scans are deemed acceptable, full texts will undergo OCR conversion using Google Cloud Vision (for all DL editions, see “Ingest into Readux2” below for details)</td>
<td>DPS</td>
</tr>
</tbody>
</table>
| 12 | Gather the appropriate files and metadata for ingest into Readux 2:  
- Positional text data files  
- TIFF files  
- Canvas spreadsheet (produced after the images are up on an image server)  
- Manifest metadata (see metadata schema)  
- Volume PDF | DPS/DPA |
| 13 | Ingest into Readux 2:  
- Convert TIFFs to JPEG  
- Put JPEGs on Image Server  
- Generate OCR Data Files  
- Upload OCR Data Files to GitHub  
- Create Manifest in Readux  
- Upload Canvas Spreadsheet to Readux  
- Upload PDF to Readux | DPS/DPA |
| 14 | QA 4: Readux  
- All pages present and in order  
- Page images should be rotated properly  
- OCR should be selectable  
- Metadata should render onscreen | DPS/DPA/PD/PM |

---

### Sample Scan Review Rubric

<table>
<thead>
<tr>
<th>Property</th>
<th>Acceptable</th>
<th>Unacceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>Lossless TIFF format or camera raw</td>
<td>JPEG, PNG, or other lossy formats</td>
</tr>
<tr>
<td>Resolution</td>
<td>600 or 400 dpi</td>
<td>Less than 400 dpi</td>
</tr>
<tr>
<td>Color</td>
<td>Full color (24-bit)</td>
<td>Black and white, grayscale</td>
</tr>
<tr>
<td></td>
<td>Accurate color</td>
<td>Noticeable distortion of original color palette.</td>
</tr>
<tr>
<td>Focus</td>
<td>Sharp focus, all content should be easily read and accessed at a high resolution</td>
<td>Blurriness, halos around letters or notes</td>
</tr>
<tr>
<td>Scratches</td>
<td>No visible scratches</td>
<td>Visible scratches</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Distortion</td>
<td>None or minimal warping/bowing/distortion of the content or page image as a whole.</td>
<td>Noticeable warping/bowing/distortion of the page image or content therein. Moire patterns (herringbone or crosshatched or dotted pattern in your scanned images)</td>
</tr>
<tr>
<td>Distractions</td>
<td>No distracting information in the image. Background of image should be black or gray. Nothing should be in the shot but the page image and the background.</td>
<td>Additional information or objects in the background of the image. Visible presence of scanning apparatus.</td>
</tr>
<tr>
<td>Obscuring</td>
<td>Content should not be obscured by the gutter of the page or digitization apparatus.</td>
<td>Images where gutter cuts off the content in the margin of the page. Content obscured by digitization clips or other pieces of the scanning device.</td>
</tr>
</tbody>
</table>
Appendix G: Visualizations of List of Digitization Candidates

This section features information about the 1,284 Sounding Spirit Digital Library digitization candidates, a subset of the “Checklist of Southern Sacred Music Imprints, 1850–1925.” The following visualizations illustrate the distribution of publication dates across the project’s temporal scope, map the list’s geographical variety, and chart the range of denominations, non-English languages, notation styles, and other characteristics of the works. The list of digitization candidates is accessible as a spreadsheet: https://docs.google.com/spreadsheets/d/1qKR_K2WWf4ak1FD5RlgcVElqripwikLpwQu2KAo71O8/. The spreadsheet version of the digitization candidates includes full bibliographic data and contextual information about the works. Each spreadsheet field is defined in the data dictionary created during the compilation of the checklist (https://doi.org/10.15139/S3/VQT4D5).

Date of Publication by Decade

The digital library will include numerous works published during each decade of the project’s temporal scope. The distribution of publication dates illustrates the decline in economic and publishing activity due to the Civil War and the expansion of publishing activity beginning in the 1880s as the southern economy rebounded. The digital library includes major editions of works published before 1850 that remained in print and in active circulation.
The digital library will include works associated with every southern state and numerous border states. States that developed robust publishing infrastructures often associated with gospel music, such as Tennessee, Georgia, Texas, and Virginia, are particularly well represented. In other areas, local publishers were rare or had limited capabilities during our temporal scope, leading many southern authors or editors to find publishers in Ohio, New York, Illinois, Pennsylvania, and Massachusetts, where publishers frequently issued books marketed toward southern audiences. The digital library also includes works from Oklahoma/Indian Territory connected to the southern United States by the Trail of Tears and from the West Coast and urban North representing the music making of populations that emigrated from the South during the Great Migration.
Music Notation Systems

The digital library includes works adopting a range of competing music notation systems in addition to the conventional "round note" system associated with western music pedagogy. Alongside various "seven-shape" methods, the corpus also includes the older "four-shape" method, systems of numeral notation, an instance of tonic sol-fa, and a work that combines round note, seven-shape, and four-shape compositions. Numerous sacred works feature hymn texts without notated music, represented by “nil.”
Although English-language texts predominate in the corpus, the digitization candidates include seventy-nine works in thirteen additional languages, including those associated with immigrant communities and Native American populations.

The digital library will include hundreds of texts associated with particular religious groups. Methodist and Baptist works were the most widespread across the South during this period. Restorationist, Presbyterian, Pentecostal, and Lutheran publishers were also highly active. The corpus includes multiple works associated with churches connected to German immigrant communities, several specifically interdenominational works, and works representing Catholic and Jewish music making.
## Characteristics Represented

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Books</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadside</td>
<td>5</td>
</tr>
<tr>
<td>Confederate Imprint</td>
<td>11</td>
</tr>
<tr>
<td>Convention Gospel</td>
<td>137</td>
</tr>
<tr>
<td>Dispersed Harmony</td>
<td>22</td>
</tr>
<tr>
<td>Distributor in South</td>
<td>48</td>
</tr>
<tr>
<td>Ethnographic</td>
<td>17</td>
</tr>
<tr>
<td>Female Author/Publisher</td>
<td>69</td>
</tr>
<tr>
<td>Fraternal Society</td>
<td>11</td>
</tr>
<tr>
<td>Gospel</td>
<td>481</td>
</tr>
<tr>
<td>Liturgical/Devotional Content</td>
<td>66</td>
</tr>
<tr>
<td>Missionary Use</td>
<td>93</td>
</tr>
<tr>
<td>Native Use</td>
<td>26</td>
</tr>
<tr>
<td>Non-White</td>
<td>167</td>
</tr>
<tr>
<td>Occasional</td>
<td>17</td>
</tr>
<tr>
<td>Out-of-Country Publisher</td>
<td>13</td>
</tr>
<tr>
<td>Out-of-Region Publisher</td>
<td>150</td>
</tr>
<tr>
<td>Revival Use</td>
<td>226</td>
</tr>
<tr>
<td>Scientific Music</td>
<td>93</td>
</tr>
<tr>
<td>Sheet Music</td>
<td>8</td>
</tr>
<tr>
<td>Singing School Use</td>
<td>149</td>
</tr>
<tr>
<td>Spirituals</td>
<td>58</td>
</tr>
<tr>
<td>Still in Use</td>
<td>13</td>
</tr>
<tr>
<td>Sunday School Use</td>
<td>421</td>
</tr>
<tr>
<td>Tract Society Publication</td>
<td>30</td>
</tr>
</tbody>
</table>

This table documents additional features of the digitization candidates including context, genre affiliation, audience, and format. See the data dictionary for the "Checklist of Southern Sacred Music Imprints, 1850–1925" for criteria and additional details ([https://doi.org/10.15139/S3/VQT4D5](https://doi.org/10.15139/S3/VQT4D5)).
Appendix H: Lists of Descriptive Entries

Preliminary Collection List

Works in the Sounding Spirit digital library will be organized into 100 overlapping collections, organized around humanities themes that offer pathways into the corpus. Each collection will feature a 300–500-word collection description that situates the included works with respect to the project’s core themes of race, place, religion, culture, and modernity.

This preliminary list of collections draws on the key themes associated with criteria for inclusion in the library and foregrounds the range of works represented in the corpus. Collections descriptions included in the pilot digital library (https://library.soundingspirit.org/collections/) are noted below.

Key

* = Corresponds to a Collection and Collection Summary published in the pilot digital library

1. Geography
   a. Alabama*
   b. Arkansas
   c. Delaware
   d. District of Columbia
   e. Florida
   f. Georgia*
   g. Kentucky*
   h. Louisiana*
   i. Maryland
   j. Mexico*
   k. Mississippi
   l. Missouri*
   m. North Carolina
   n. Ohio*
   o. Oklahoma/Indian Territory*
   p. Pennsylvania
   q. South Carolina
   r. Tennessee*
   s. Texas*
   t. Virginia*
   u. West Virginia

2. Sacred
   a. Adventist
   b. Assemblies of God
   c. Baptist*
   d. Brethren
   e. Catholic
   f. Church of Christ
   g. Church of God (Anderson)
   h. Disciples of Christ
   i. Episcopalian
   j. Full Gospel
k. German Evangelical
l. Holiness
m. Latter-Day Saints
n. Lutheran
o. Mennonite
p. Methodist*
q. Moravian
r. Nazarene*
s. New Thought
t. Orthodox
u. Pentecostal
v. Presbyterian*
w. Quaker
x. Restorationist*
y. Shaker
z. Spiritualist
aa. Universalist
bb. Jewish
c. Interdenominational
dd. Liturgical/devotional

3. Style
   a. Dispersed harmony
   b. Gospel
   c. Scientific music
d. Spirituals
e. Lined-out hymnody

4. Notation
   a. Four-shape
   b. Numeral
c. Seven-shape

5. Context/Use
   a. Convention gospel
   b. Denominational publications
c. Missionary hymnody
d. Occasional music
e. Revival hymnody*
f. Singing school music
g. Sunday school hymnody*
h. Still in use

6. Format
   a. Annuals/Serials*
b. Broadsides
c. Ethnographic/scholarly works
d. Sheet music
e. Text-only hymnals
f. Tunebooks
g. Tracts

7. Race/Ethnicity/Nation
   a. African American*
b. Cherokee
c. Choctaw
d. Creek*
e. Delaware
f. Hispanic
g. Potawatomi
Preliminary Bibliographic Essay Topics

Sounding Spirit will recruit an interdisciplinary group of fifteen leading humanities scholars to write bibliographic essays that contextualize the digital library’s works and their interventions around key humanities themes. Foregrounding Sounding Spirit’s emphasis on race, place, religion, and culture, these bibliographic essays will invite and model scholarly engagement with the corpus and promote the library’s contents to a wide-ranging audience. The following fifteen sample essay topics are organized around terms central to the library’s priorities and organization.

1. Race
2. Geography
3. Religion
4. Identity
5. Genre
6. Format
7. Language
8. Modernity
9. Gender
10. Notation
11. Pedagogy
12. Vernacular
13. Circulation
14. Text
15. Audience