TISMIR

Community Based Music Information Retrieval: A Case Study of Digitizing Historical Klezmer Manuscripts from Kyiv

RESEARCH

YONATAN MALIN
CHRISTINA CROWDER
CLARA BYOM
DANIEL SHANAHAN

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

In this article we provide a case study in the datafication of historical handwritten manuscripts, which diversifies the repertoire, approaches, demographics, and institutional partnerships of MIR. The Kiselgof-Makonovetsky Digital Manuscript Project (KMDMP) is a community-based project to digitize music and text, teach, and make music from facsimiles of manuscripts held by the Vernadsky National Library of Ukraine. The corpus comprises 850 high-resolution photographs of handwritten music manuscripts and catalog pages, with a total of around 1,300 melodies. Much of the music was collected by pioneering Belarusian ethnographer Zusman Kiselgof among Jewish communities in the 'Pale of Settlement' (mostly in modern Ukraine and Belarus) during the An-Ski Expeditions of 1912–1914. The repertoire is mixed, combining typical Jewish dance and non-dance genres, European society and folklore dance music, and a relatively small quantity of songs and liturgical chant settings. The project simultaneously encodes music in formats accessible to computational musicology and enhances a creative musical community and deeply valued heritage. We introduce the project in dialogue with a recent article by Georgina Born on diversity in the field of MIR; present the material, issues for datafication, and results thus far; describe project elements that enhance musical community; demonstrate the diversity of participants with respect to age, gender, nationality, and profession; outline implications for MIR and computational ethnomusicology; and suggest new funding models and partnerships in support of cultural heritage documentation, preservation, continuity, and analysis.

CORRESPONDING AUTHOR:

Yonatan Malin

University of Colorado Boulder, US

yonatan.malin@colorado.edu

KEYWORDS:

heritage music; klezmer; Jewish music; crowdsourcing; computational ethnomusicology

TO CITE THIS ARTICLE:

Malin, Y., Crowder, C., Byom, C., & Shanahan, D. (2022). Community Based Music Information Retrieval: A Case Study of Digitizing Historical Klezmer Manuscripts from Kyiv. *Transactions of the International Society for Music Information Retrieval,* 5(1), 208–221. DOI: https://doi.org/10.5334/tismir.135

1 INTRODUCTION

In a recent article on diversity in the field of Music Information Retrieval (MIR), Georgina Born asks,

> Which masters or mistresses does MIR serve—the profit-seeking imperatives of commercial music tech corporations and online music services, entangled as they are in the recorded music industries? And which mistresses *should* MIR serve in order to diversify its goals, partners and worldly effects? In sum: could MIR cultivate a more plural set of orientations and institutional partners so as to include non-commercial, publicly-oriented initiatives aimed at enhancing human musical flourishing? (Born, 2020: 196)

Our paper offers a case study of one recent and ongoing MIR project that is tied to a 'non-commercial, publicly oriented initiative aimed at enhancing human musical flourishing'. The project simultaneously encodes music in formats accessible to computational musicology and enhances a creative musical community and deeply valued heritage. While the impetus for the project—the desire to make a new corpus available to klezmer practitioners, as well as researchers—is rather different from the purpose of many MIR projects, the methods and lessons learned may be of value for diversity-related initiatives in MIR. This includes initiatives that address the four dimensions of diversity identified by Born (2020): demographics, repertoire, foundational principles, and the real-world institutions and communities that MIR serves.

Our case study focuses on the Kiselgof-Makonovetsky Digital Manuscript Project (KMDMP), a community-based project initiated in fall 2020 to digitize music and text, and to teach and make music from a collection of handwritten manuscripts held by the Vernadsky National Library of Ukraine. Much of the music was collected by pioneering Belarusian ethnographer Zusman Kiselgof among Jewish communities in the 'Pale of Settlement' (mostly in modern Ukraine and Belarus) during the An-Ski Expeditions of 1912–1914.[1] The corpus also includes a manuscript of music for Jewish weddings notated by the violinist and band leader Avram Yehoshua Makonovetsky (1872–?).[2] Between these two sources, the corpus contains approximately 1,300 melodies. These are Jewish and other Eastern European melodies in a genre now known as klezmer: music performed by professional guild members (pl. *klezmorim*) for Jewish weddings.

This collection is a snapshot in klezmer history that demonstrates musical qualities that persist today—the mixing of modality and tonality, variety of functions (dance, ritual, listening), and influence of neighboring musical traditions on the klezmer repertoire. Most of the music is notated as single-line melodies, which were typically played on violin and sometimes on clarinet, flute, trumpet, or other melody instrument. The main melody instrument was usually accompanied by one or more additional instruments, creating flexible ensembles ranging from two to twelve or more members depending on the occasion (e.g., violin and tsimbl duo up to violin-led wind ensemble). Though it is known that music notation reading and writing was part of the education of klezmer musicians (Lipaev, 1904), we understand these scores to be documents of musicians' repertoires rather than scores for performance. Accompaniment style, ornamentation, phrasing, and harmonization (not included in the manuscript pages) were and continue to be learned orally, and tunes were played from memory.[3]

Structural choices made for the project were influenced by the anticipated needs of a larger digital humanities project—The Klezmer Archive. The Klezmer Archive Project is currently in a research and development stage, aiming to create a 'universally accessible, useful resource for interaction, discovery, and research on available information about klezmer music' (Klezmer Institute, 2019). It will bridge the gap between oral history and archives by being a space where culture bearers and community members can engage with archival items and one another by drawing connections between tunes, discussing genre classifications, searching for unnamed melodies, and comparing recorded versions of the same tune. The KMDMP is a testing ground for methods and tools to be used in the broader Klezmer Archive. The Klezmer Archive Project is funded by a Phase I Digital Humanities Access Grant from the National Endowment for the Humanities for 2021–2022. As a case study for the Klezmer Archive, small parts of KMDMP work have been supported by the NEH grant. KMDMP does not have independent funding and is supported solely by individual donations.

KMDMP involves the encoding of music in digital formats and technological choices that maximize community engagement on a limited budget.[4] Some of the choices in turn highlight challenges of current technologies (e.g., lossiness in music XML between notation programs) that could be solved by further work in MIR. The population engaged with the project is diverse with respect to gender, sexual orientation, profession, age, and nationality. It is our hope that the project and its strategic deployment of existing technologies offers models for achieving a degree of demographic, stylistic, and conceptual diversity that MIR aspires to.

We would like to emphasize the cultural significance of the KMDMP at the outset. The culture of Ashkenazic (Eastern European) Jewry was decimated in the 1930s through the 1950s by the Holocaust and by Stalinist repression, which included the dissolution of cultural organizations and the murder of many of their leaders. A branch of Ashkenazic culture including the language (Yiddish) and music (now called klezmer) had been transplanted to North America by Jewish immigrants circa 1880–1920, and it evolved for a few decades in the

American context. After World War II, American Jews were less interested in the language or music of their parents or grandparents, gravitating instead to jazz and American folk music, or to the Hebrew songs and new Jewish identity of the State of Israel. In the 1970s and '80s, when members of the following generation became interested again in the music of their ancestors, they had few resources to work from. The available resources have been growing since the 1980s even as surviving culture bearers from the early twentieth century have passed away. These include 78 RPM records, ethnographic recordings on wax cylinders (some now digitized), and the scholarship and notated collections of the pre-eminent Ukrainian Jewish ethnomusicologist Moshe Beregovski (1892–1961; see Beregovski, 1982; 2015). With all of this in mind, the manuscripts in KMDMP represent a fivefold increase in the number of klezmer tunes in historical notation from the Eastern European repertoire.[5] The digitized manuscripts and their dissemination are also important now, in 2022, due to the threats to cultural heritage in the Russian war against Ukraine.

The corpus is significant both within and beyond the bounds of Jewish culture. Eastern European Jewish musicians or *klezmorim* frequently played with non-Jewish musicians and for non-Jewish audiences. The repertoire is eclectic, including typically 'Jewish' genres such as *freylekhs* and *skotshnes*, 'transitional' repertoire such as *bulgars*, 'co-territorial' repertoire such as *mazurkes* and *kozatshoks*, and broader European repertoires such as polkas, marches, waltzes, and more (Feldman, 2016, chapter 7; Mishiro, 2021). Since this repertoire was collected among Jewish musicians, it serves as a valuable source for studying the interaction among folk music subcultures in Eastern Europe.

Once fully digitized, this corpus will add to the diversity of corpora available for computational ethnomusicology, building on work by Xavier Serra (2011; 2014) and the CompMusic project (CompMusic, 2022). The CompMusic project, however, focuses primarily on audio collections, whereas KMDMP involves symbolic data in the form of digitized scores. The CompMusic project also focuses on art music traditions (Serra, 2014), whereas the KMDMP documents a tradition that is professional but not typically understood as 'art music'. In this sense, this corpus is more aligned with much of the work utilizing symbolic data as a representation of folk tradition, such as the Densmore Collection of Native American Folk Songs (Shanahan and Shanahan, 2014), as well as the Meertens Tune Collection (van Kranenburg et al. 2016), and the Essen Folksong Collection (Schaffrath and Huron, 1995).

Serra (2014: 2–3) outlines five criteria for the design of research corpora: purpose, coverage, completeness, quality, and reusability. In the KMDMP case, the project is not to design a corpus as such, but rather to convert a historical corpus into machine-readable formats as well as formats that are accessible to creative, performing musicians. The main purpose is not to answer a specific research question in MIR per se, but historical understanding and access. Machine-readable formats will allow musicians and researchers to search for particular tunes or sets of tunes based on a variety of criteria, and to develop other forms of computational analysis. Ilyefalvi (2018: 386–390) describes similar folklore digitization projects for historical collections, and Gjerdingen and Bourne (2015: 5.3.1) describe using historical collections as an alternative to modern sampling.

We may nonetheless consider Serra's additional criteria in this context. For coverage and completeness, this is the broadest available collection from the early twentieth century and it is the earliest collection of its kind (Sholokhova, 2006). The quality of the historical documents is mixed: some manuscripts are hard to read, and we lack additional information on sources and uses of the material. Finally, with regard to reusability, permission to distribute the manuscripts for use by scholars and musicians (not for commercial distribution) was granted at the time of acquisition, and derivative works including digitally engraved music are governed by the Creative Commons CC-BY License with attribution to the Institute of Manuscripts of the Vernadsky National Library of Ukraine. Volunteer-contributed, digitally encoded scores are available in PDF format for immediate use (Klezmer Institute, 2022). Encoded scores will be edited and proofread before being posted in machine readable formats via the Klezmer Archive Project's GitHub or another open repository such as Zenodo. (See Section 5 for more details.)

In the following sections, we will discuss the nature of the KMDMP manuscripts and challenges of digitization (Section 2), methods for community engagement including technological choices and results to date (Section 3), information about project participants and their interests (Section 4), implications for MIR and computational ethnomusicology (Section 5), and reflections on the case study in the larger context of MIR and diversity (Section 6). A supplementary appendix provides biographical vignettes based on interviews with four project participants.[6]

2 THE KMDMP MANUSCRIPTS

The KMDMP corpus comprises 850 high-resolution photographs of handwritten music manuscripts and catalog pages in three parts:

1. 26 non-consecutive numbered notebooks (*heftn*) with handwritten music from various klezmer sources collected by Kiselgof during the An-ski Expeditions
2. Moshe Beregovski's catalog of this material
3. A 236-page music manuscript prepared by Avrom-Yeshieh Makonovetsky for Moshe Beregovski ca. 1938.

The physical notebooks, catalog, and manuscript are housed in Archive #190 at the Institute of Manuscripts of the Vernadsky National Library of Ukraine. The vast majority of the melodies are instrumental tunes, with a small but significant number of songs, *niggunim* (wordless melodies connected with spiritual intention), and liturgical pieces. The Makonovetsky manuscript has 278 items which are numbered in the source and clearly delineated. The section of Beregovski's catalog that we have (pages 68–156) lists 1,059 items; 118 of those are in notebooks that we do not have, leaving 941.[7] The notebooks that we do have include additional items not cataloged by Beregovski; these are either items that he seems to have identified as of no interest to his own work (e.g., Western social dances) or incomplete fragments. We estimate that there are between 62 and 75 melodies that do not have catalog numbers (whether they are fragments or complete items is not always clear). Thus, with items from Makonovetski (278), items listed by Beregovski (941) and additional items (circa 62–75), the corpus has approximately 1,281–1,294 items.

The music contains a mix of genre types from across the spectrum of the repertoire of a professional klezmer musician during the late nineteenth and early twentieth centuries (Lipaev, 1904). Notations range from simple two-part "AA BB" dance tunes to multi-page solo show pieces with known composers. The corpus contains many non-metric pieces for both voice (liturgical pieces) and violin (ritual wedding genres). The Kiselgof Hefts include material contributed from a number of named musicians, but also loose manuscript sheets and sheafs of *alte klezmerishe noten* (old klezmer notations). Without further information it cannot be clearly established whether notations ostensibly from one musician were written by that individual or transcribed from another manuscript or a live performance. By all appearances, the notations were written by different klezmorim for their own use, and the vast majority could be considered 'lead sheets' without an indication of intended instrumentation and harmonization. All of this presents obvious challenges to context and interpretation, and many different handwritings also make it difficult to make inferences about authorial intent in note placement and interpretation of various music symbols.

The Makonovetsky Wedding manuscript was written by a single musician in relatively clear notation with bowing and other markings for violin. The manuscript also includes precious commentary about performance contexts in Yiddish script with distinctive spellings and a stream-of-consciousness style. Transcribing and interpreting these text annotations requires deep domain expertise in pre-revolutionary Russian and Yiddish prior to YIVO standardization, as well as Jewish cultural and musical knowledge.

2.1 DIGITIZING THE MANUSCRIPTS

Alongside, or despite advances in OMR, the digitization of manuscripts like those in the KMDMP collection must be done by hand. A blend of human and machine-based input is always possible, but handwritten manuscripts by multiple authors, many of whom were writing for their own personal use, are particularly challenging for OMR. In this section, we highlight the complexity and nuance of the material that is being digitized, even when the tunes themselves are simple. Capturing this complexity is necessary, we argue, for any subsequent analysis to be historically, culturally, and musically informed.

Figure 1 shows a section of a page from Heft 37 with two tunes.[8] Numbers show textual annotations and

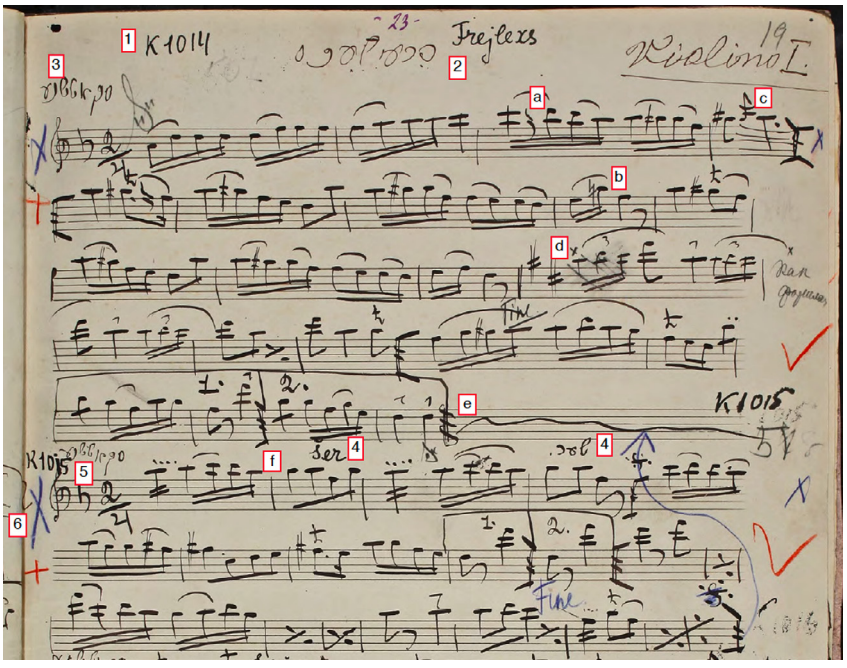

Figure 1 Sample manuscript page from heft 37.

letters show aspects of the music notation that we will discuss. First, the textual annotations indexed by number:

1. 'K1014' is Beregovski's catalog number. The tune below is marked 'K1015'.
2. The tune is identified by genre as 'Freylexs', which is a Russian spelling for YIVO standard *freylekhs*. The handwriting appears to be Beregovski's; it is similar to the writing in his catalog. The same genre designation is given in Yiddish in different ink and another hand.
3. The tune is also identified in Yiddish as a *skotshne*. This may be the designation made by the original musician, Motl Reyder. The genre designations are fluid; musicians used different genre identifiers for the same tune. This corpus can help us clarify who uses which genre designation and in what language or context.
4. The next tune is identified in Yiddish and transliteration as a *sher*, in what appears to be Beregovski's handwriting.
5. It is also identified in Yiddish as a *skotshne*. The *skotshne* markings in this and the previous tune appear to be in the same hand.
6. The number '2', barely discernible in the crease of the volume, indicates that it is the second in a series of *skotshnes*. The one after (not shown here) is numbered '3'.

And now aspects of the music notation, indexed by letter:

a. This is a natural sign on the B. Understanding of the mode, which is identified among klezmer musicians as either altered Dorian or *misheberakh* helps us read this.
b. This is a natural sign on G, more in line with modern usage. It appears to be in a different hand and may be a later addition.
c. Klezmer musicians would often read the 'grace note' as a *krekhts*, a common form of klezmer ornamentation that mimics an emotional break in the voice. A likely pitch for a *krekhts* here would be B above the A.
d. There is a triple sixteenth run up to the D which appears to be crossed out. An asterisk above them points us to the text annotation on the right. In Russian, it reads 'Как форшлаг' (like a grace note).
e. This symbol is a *serpent segno* that functions in the same way as a *dal segno* (D.S.) marking.
f. In the beginning of bar 2 of the sher, notes above the staff are not distinguished by height. The notes are G and A. The G has the stem on its left. The A has the stem coming down from the middle of the notehead, which is longer horizontally and functions simultaneously as a ledger line. We see the same effect in measure 1; the four sixteenths are <A-C-Bb-A>.

Figure 2 shows the engraved score with annotations in the format that project participants are asked to use. See Sections 3 and 5 for a further discussion of data formats.

Cultural knowledge is brought to bear in the further identification of tunes. Klezmer clarinetist and project participant Hannah Ochner observed that the tune of K1015 was also recorded as 'Alle in Einem' by the Abe Schwartz Orchestra (1920) with a slight variation (F# in bars 2 and 4). Connections like these are being documented for future reference.

In the first 18 months of the project approximately 1,114 of the 1,281–1,294 items in the corpus have been digitally notated. Nearly all of the text annotations have been transliterated and translated, setting the project up to transition into an editorial phase.

3 PROJECT DESIGN FOR COMMUNITY ENGAGEMENT AND OUTREACH

The technique of crowdsourcing is widely used to harness the interest of a large group of participants in research.[9] Crowdsourcing makes the research process more transparent, but there is often still a hierarchy with researchers at the top. For instance, the webpage for Zooniverse, a widely used platform for crowdsourcing, states: 'This research is made possible by volunteers–more than a million people around the world who come together *to assist professional researchers*' (Zooniverse, n.d.; emphasis added). What if instead, we understand the material as being of equal importance for researchers and musicians or artists, and those who are both? The KMDMP was designed to reflect these values.[10]

Anna Rogers, who traveled to Kyiv and obtained the manuscript images in the summer of 2017, articulated a philosophy that continues to guide the project: a desire to release all the material to everyone in the community as the first step—without curation, commentary, or restrictions.

> The klezmorim interviewed by Kiselgof were professionals, with an eclectic collection of different tunes in their repertoire that let them earn a living, generation after generation, adapting to the new tastes and audiences. Their voices faded through numerous social cataclysms, but luckily, the archive survived. *While these materials are very important for research, their biggest impact is going to be when they return to the fingers of the musicians and the feet of the dancers, when they continue their journey through human hearts*. Luckily, we discovered this time capsule at a post-revival time, when there is a generation of new klezmorim more than capable of processing and reviving this repertoire. The communities of Yiddish Summer Weimar, Yiddish

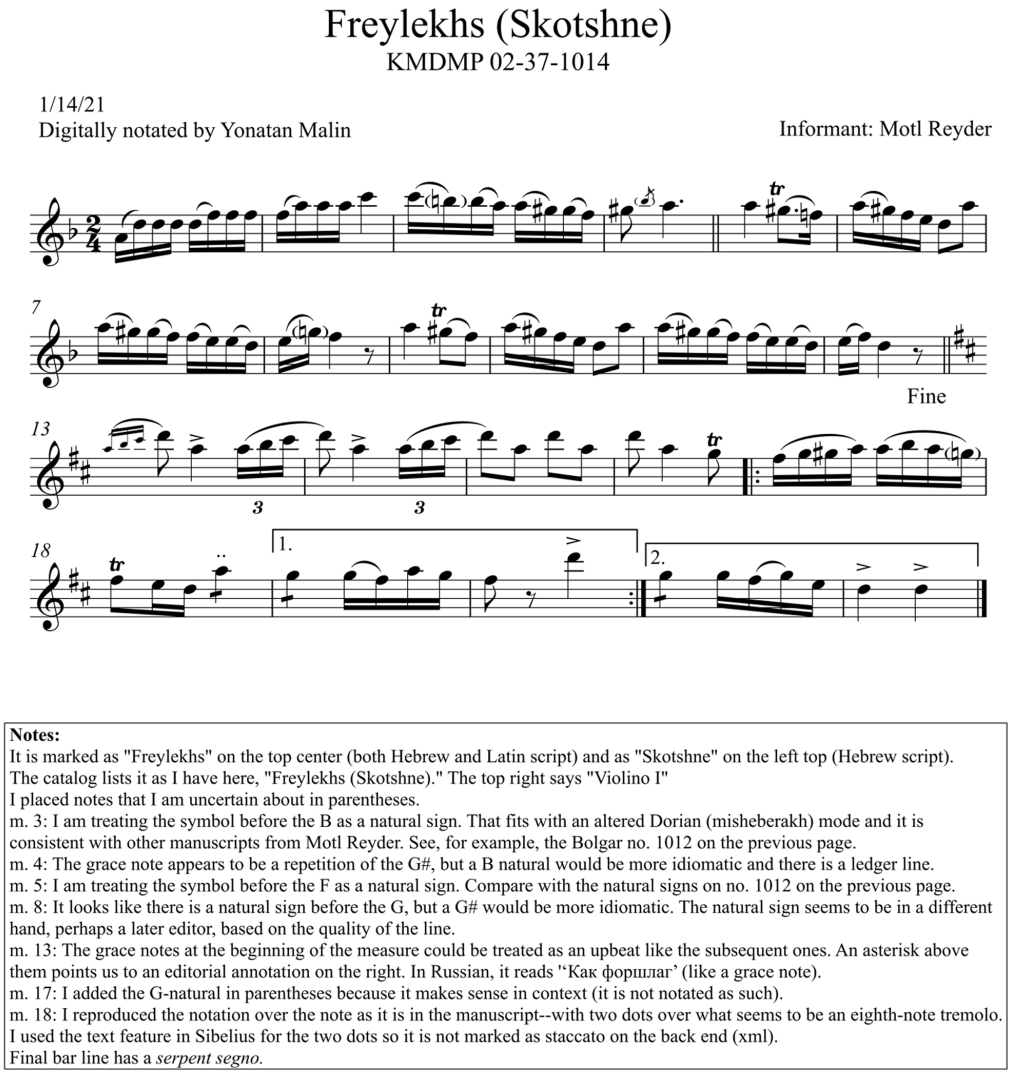

Figure 2 Engraved score with annotations for the tune in Figure 1.

> New York, KlezKanada, etc. are the real heirs to those who contributed to the archive in the first place. Why not pass them their inheritance, and let them develop it further—while helping to create a curated academic edition? (personal communication with Crowder, emphasis added)

With this quote in mind, we can return to Georgina Born's discussion of MIR's ontological assumptions. From the outset, we understand the KMDMP manuscripts not primarily as forms of musical data to be mined for knowledge, but as historical vessels that bridge trauma and carry the potential for cultural revitalization. The ontology is thus one that 'exceed[s] knowledge and representation by taking into account the embodied, social and material aspects' of the musical culture (Born, 2020: 199).

As Rogers observes, the material's primary significance centers around what current klezmer musicians can do with it, building bridges to the past with new creative endeavors. Given the desire to serve today's klezmer community, and acknowledging the limitations of a young organization launching an unfunded initiative, the project infrastructure was created to support its guiding principles:

- Free access to materials.
- Reduce/eliminate barriers to participation.
- Make digital outputs immediately available.
- Practice open-source principles.

To make material available for project launch the images were: 1) organized into PDFs by heft, 2) digitally 'stamped'

with CC-BY 4.0 attribution and logo; and, 3) stored in an accessible place. Christina Crowder, the project director, took on these tasks, developing numbering systems, PDF design, logo, and a public-facing web page. Source PDFs are accessed via Google Drive folder and a Google registration form grants access to collaboration resources. Participants access the project from a password protected web page. In addition to announcements, the page provides a collection of links to the 'KMDMP Commons' folder containing guides and how-tos along with user-generated sub-folders. Distributed work on the corpus is coordinated via Google sheets.

KMDMP relies on volunteer encoding and does not require participants to purchase notation software. In order to keep the barriers for participation as minimal as possible, the project team decided to be as platform neutral as possible with regard to music notation software. Participants with previous notation experience often use Sibelius, Finale, or Dorico. Others, particularly individuals new to digital notation, use the free program MuseScore. Zooniverse was considered and rejected because it would have required custom training in a text-based music notation format such as Kern (Huron, 1997), which is not intuitive for most musicians. There are also no musical equivalents to the crowdsourcing manuscript transcription software FromThePage (Brumfield and Brumfield, 2022).

Allowing for this flexibility now means that editorial teams in the future will have to deal with lossiness during file conversions between music XML generated in different programs and human error in the encoding process, but it was decided early on that it is easier to edit scores that exist than to start from scratch. Participants are asked to produce three files for each notated score: native format, uncompressed music XML, and PDF. Notators are asked to observe the original manuscript closely, and to digitally reproduce as best they can what they see on the manuscript page. The collection of community-digitized material will be edited by a smaller team.

Simple automations can help with community-based digitization projects. In the first few months, participant onboarding and digitized score processing was done by hand. As this became unwieldy, Crowder and Klezmer Archive team member Max Rothman developed automations for these processes using the low-cost business automation program Integromat, now rebranded as 'Make' (Make, 2022). The *Onboarding Scenario* is triggered by a new member questionnaire. It sends an introductory email, gives permission to Google sheets and folders, adds the participant to email lists based on their choices, and flags them for an orientation over Zoom. Figure 3 shows one branch of the *Score Processing Scenario,* which contains 72 individual processing modules. It standardizes file names, routes files to the appropriate places based on the tune (indexed by KMDMP number), format (.PDF, .XML, .NATIVE), type (with or without text), and version (new/updated). Version control on updated score submissions is managed through the scenario, which attaches new versions to the original submission via Google Drive, preserving all iterations. The Score Processing Scenario also sends confirmation emails to the uploader and project administrators, and error notifications to a dedicated Slack channel.

Providing the technical infrastructure to support a community-based project is one task but engaging and sustaining participation is another. Project organizers recognized the need to build on the initial enthusiasm and bring the community together through virtual events. The first KMDMP Digitizathon—a 30-hour marathon held via Zoom in January 2021—featured dozens of musicians, researchers, and linguists who agreed to host digitization and translation sessions, and conversations about issues surrounding the digitization process. What do you do when you don't know what a particular symbol means? What if the piece is in 3, but there is a bar in 4? What if I can't tell if they meant an F# or an F-natural in measure 22? I think they forgot a repeat sign here – what do I do?

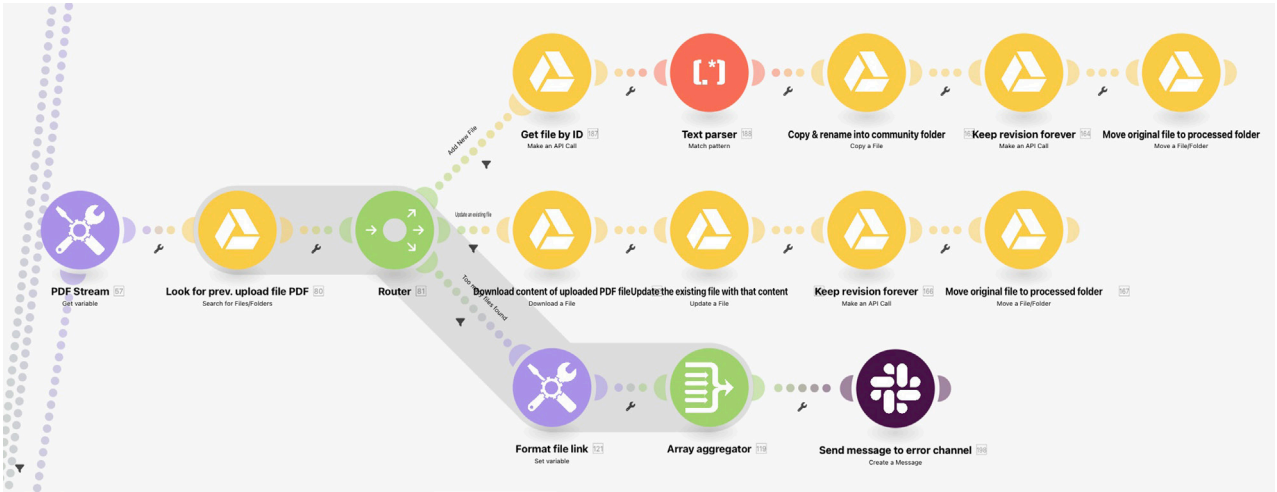

Figure 3 Integromat Score Processing Scenario showing a portion of the PDF file stream.

What is the goal for this particular phase of the project? What do the project participants want from the project?

Based on feedback after the first event, the administrators and motivated project members worked together to adapt and add components to meet the needs of the community. The digitizathons continue to be productive, invigorating gatherings where experienced specialists and learners at many levels interact in the 'scavenger hunt' or 'puzzle' of the corpus together. Digitization also provides participants with a sense of ownership and intimate knowledge of the tunes they work on. These non-hierarchical, active learning environments have provided an opportunity for many individuals to step into leadership roles and become champions for the project.

4 PROJECT PARTICIPANTS AND LEADERSHIP

When conceiving of and implementing the project, KMDMP leadership was focused on creating a space that welcomed individuals with a variety of technical expertise, musical backgrounds, and academic experience. Choices were made to foster this inclusion—platform agnosticism and non-hierarchical learning spaces as described above. Oudshoorn, Rommes and Stienstra (2004, quoted in Born, 2020: 196) observe how an 'I-methodology' that universalizes engineers' predominantly male subjective experience with technology can limit the diversity of users. Our approach has sought to allow for diversity in technological experience and inclination. Project participants have been diverse with respect to gender and age, though less so with respect to race, ethnicity or socio-economic background. Participants also include those who identify as Jewish and non-Jewish, and both religious and secular.

As of July 2022, there are 229 registered project participants. Registration data indicates the diversity of languages and hence international reach of the project. The languages in alphabetical order are Arabic, Aramaic, Dutch, English, French, German, Greek, Hebrew (21 speakers), Hungarian, Italian, Japanese, Ladino, Latvian, Lithuanian, Polish, Portuguese, Russian (37 speakers), Spanish, Ukrainian, and Yiddish (103 speakers).

KMDMP has been led and supported by women at all stages of the project. While open to all, the most engaged participants and the project organizers are women. After the first Digitizathon event, Jutta Bogen commented, 'I am pleased that the implementation of the project bears an unpretentious, feminine signature' (personal communication with Crowder). This engagement is felt both in 'work product' as measured in the overwhelming majority of notated tunes, text transcriptions, recordings, and review of melodies by women participants, but also through less quantifiable ways. Crowder observes that there is a great deal of spoken and unspoken excitement about engaging with repertoire that has not already been occupied by others in what remains a musical space where power and influence flow primarily to male musicians and scholars.[11] There is a palpable sense of opportunity for women musicians to make their mark with this repertoire and to create what may become canonical interpretations.

While building for technical, musical, and academic diversity may not necessarily indicate that individuals of all ages are welcome, the KMDMP community also demonstrates that technical skills can be acquired by individuals of all ages. Several older klezmer enthusiasts, such as Judit Nemtanu Shapiro (Bordeaux, France) and Dan Berkowitz (Moore, Oklahoma, USA) have learned how to use MuseScore to more fully participate. While some of the klezmer professionals from the revival generation have not contributed digitized scores, many generously share their expertise at digitizathons and other fora. Likewise, individuals of all ages have provided technical support for the group. It is also notable that substantial efforts to translate text in the manuscripts are being led by individuals under the age of 40, folks like Eléonore Biezunski (Sound Archivist at YIVO Institute for Jewish Research, New York, NY), Hannah Ochner (Stuttgart, Germany), Mariko Mishiro (Tokyo, Japan), and Asya Vaisman Schulman (Yiddish Book Center, Amherst) among others.

Project leaders and participants have musicological and personal interests that could be addressed with research in MIR and computational ethnomusicology. Understanding the geographic diversity of the repertoire in terms of its musical complexity, rhythmic density, and other factors (both between the northern and southern repertoires of eastern Europe, and between European and North American repertoires) is of great interest to the project leaders. Participants shared additional interests in interviews completed in the spring of 2022. Hannah Ochner (a physicist from Stuttgart, Germany) is interested in tune variants, the variable use of genre names, and historical information about the Jewish wedding. Jordan Hirsch (a professional trumpet player from Teanick, New Jersey, USA) is interested in the fluidity between Hasidic music (from the Hasidic branch of Judaism founded in the 18th century) and repertoire that is more commonly associated today with klezmer. Jutta Bogen (a violinist and violin teacher from Konstanz, Germany) loves working closely with the manuscripts; she also combines KMDMP repertoire and folk music from the Black Forest region in Southwest Germany in adult education classes. The personal and musical journeys of Daniel Berkowitz (a meteorologist from Moore, Oklahoma, USA) illustrate how music and social networks are intertwined. The Appendix provides interview dates and biographical vignettes on these four participants.[12]

The project is ostensibly about rendering music and text to computer-readable notation, but the social relations (past and present), memory, and history embedded in the notation are of equal importance to the KMDMP community. In KMDMP, the often solitary act of

digitizing paradoxically supports and strengthens cultural memory and social connection. As Jordan Hirsch put it,

> The KMDMP project has brought people together from all over the world, working together … Ah, what a gift, what a gift, I love it, I love it … People have become friends, virtually, at a time when everyone is feeling beleaguered and besieged. And they have done it over something that is already infused with so much love. (Interview, February 10, 2022)

The larger Klezmer Archive project will document networks of musical and personal connections along with musical items using semantic knowledge graph technologies.

5 IMPLICATIONS FOR MIR AND COMPUTATIONAL ETHNOMUSICOLOGY

The digitization of the KMDMP corpus will enable broad corpus and computational analyses, furthering not only our understanding of klezmer and Eastern European folk musics but also the fields of MIR and computational ethnomusicology. While much of the work in MIR deals with audio files, symbolic notation also has an important role to play. Areas of investigation such as tonic detection, identification of melodic motifs and characteristic rhythmic patterns, and automatic genre identification can be pursued with symbolic notation as with audio files. Tzanetakis (2014) provides an overview of these and other areas of investigation in computational ethnomusicology (see also Gómez et al., 2013).

KMDMP data management and storage will be administered by the Klezmer Archive project, a project that prioritizes the principles of data management and stewardship presented by Wilkinson et al. (2016), indicating that data should be findable, accessible, interoperable, and reusable (FAIR). To ensure the quality of the data set, the volunteer contributed scores will be edited by an appropriately compensated editorial team before being made available to the public via the Klezmer Archive Project GitHub or another open repository such as Zenodo.[13] This process will begin with a test corpus of 100 representative tunes for use by the Klezmer Archive Project and player folios for use by the community. The data will be made available under a Creative Commons License in standard music encoding formats such as MusicXML, MEI, and Kern for use by the MIR community, computational ethnomusicologists, and independent researchers. Additionally, the Klezmer Archive will make the data accessible to non-technologists via a user-friendly online interface that will facilitate search and browsing and allow for downloading. The data will be available throughout and a minimum of five years beyond the lifetime of the Klezmer Archive Project.

The KMDMP corpus will be particularly valuable for the study of oral transmission and tune families since some of the melodies are variants of each other and of tunes from other sources. This would build on research on tune families in the Meertens Collection, summarized by van Kranenburg et al. (2019: 19–21). The computational study of tune families draws on the knowledge of experts in particular traditions: musicians, scholars, and aficionados who are immersed in the repertoire and can identify the relevant families. The community of klezmer musicians is promising in this regard because community members are actively engaged in the music and in discussions about it. Facebook groups such as the Beregovski Online Forum (625 members) pools the knowledge of practitioners around the world, and the Klezmer Archive Project is creating 'a space where culture bearers and community members' can draw connections between tunes, discuss genre classifications, compare recorded versions of the same tune, and engage in other ways (Klezmer Institute, 2019). All of this will allow for research that combines community knowledge and computational analysis.

The KMDMP corpus is part of the larger Klezmer Archive Project as mentioned above (Section 1), and this larger project will itself be innovative with respect to library science and digital archives. The project will not only make musical and historical metadata more accessible to people and machines but will also elevate the valuable cultural context. Metadata will be structured in four levels: (1) traditional catalog information adapted for a transnational, multilingual oral tradition; (2) domain-specific metadata such as form, genre, social/ritual function, mode, and rhythmic patterns; (3) semantic metadata that documents human and musical relationships, with the relationships themselves also becoming items or 'nodes', and; (4) commentary and discussion of the kind shared in scholarly publications and community spaces (e.g., Yiddish culture festivals, workshops, and Facebook groups). The Klezmer Archive thus accounts for music as a 'total social fact' (Holzapfel et al., 2018), and it will allow for further investigations of the social nature of informationally enriched music.

With the KMDMP corpus we will be able to further research at the intersection between symbolic and audio MIR. The metadata for individual KMDMP items will include information about recordings from other sources when they exist. We will then be able to build on work by Sentürk et al. (2014) that links scores and audio recordings in a predominantly oral tradition, where performers vary the melodic lines and change structural aspects (e.g., section repetitions). As Sentürk et al. observe, 'In general, approaches of audio-score alignment assumes that the score and the target audio recording are structurally identical' (2014: 35). This is not the case in klezmer music, in the Turkish *makam* tradition that Sentürk et al. study, or in many other traditions outside of European

classical music. Additionally, we hope that this provides resources for future work in optical music recognition. As there is currently a dearth of resources using handwritten manuscripts that contain ground truth, we hope that this resource can provide exactly that.

Gómez et al. (2013: 111) observe that the set of papers on computational ethnomusicology in the special issue of the *Journal of New Music Research* hardly touch on core issues in the field of ethnomusicology. Serra (2011, 151) similarly advocates for combining approaches from MIR and musicology. There is more recent work that addresses core ethnomusicological issues such as genre, mode, and oral tradition; see for instance case studies outlined in van Kranenburg et al. (2019: 18–20). The KMDMP corpus offers opportunities to continue this work.

The topic of musical modes (including Arabic *maqām*, South Asian *rāga*, South-East Asian *pathet*, Jewish *shteyger*, and others) has been important in ethnomusicology. The recent symposium *Rethinking Musical Mode* (Kunstuniversität Graz, 2021), for instance, builds on the seminal article by Powers et al. (2001). In the field of Jewish music studies, scholars have investigated the history of discourse on mode (Seroussi 2009), modes in klezmer music (Beregovski 2015; Feldman, 2016; Horowitz, 1993; Rubin, 2020), and modes in cantorial music and everyday synagogue chant (e.g., Frigyesi, 1993; 2003; Tarsi, 2017; 2020). Jewish modes are also taught in cantorial schools (Bernard, 2005; Cohen, 2009).

Malin and Shanahan (2022) offer the first computational analysis of mode in klezmer music, and a model for the kind of work that might be done with the digitized KMDMP corpus. They work with a corpus of 254 melodies collected in the 1930s by the Ukrainian Jewish ethnomusicologist Moshe Beregovski (2015). They use pitch histograms, transition probabilities, and directed computational searches to move beyond common scalar representations of the modes.[14] The digitized KMDMP corpus will allow researchers to gather further empirical data. It will also enable cross-cultural comparisons, showing for instance how scales may be shared among Eastern European traditions while melodic tendencies and idioms differ. Musicians working at the intersection of these traditions are aware of the similarities and differences; data will provide empirical evidence confirming (or perhaps refining) their musical intuitions. In addition, a wider awareness of klezmer modes should lead music notation software developers to include key signatures associated with the modes.[15]

With all of this in mind, perhaps the most significant contribution of the KMDMP corpus for MIR will be the diversification of repertoire and cultural and social contexts. This is a repertoire that blends tonality and modality; that includes genres for dance, listening, procession, and ritual; that is performed on a wide variety of instruments from violin and clarinet to accordion, tsimbl, and trombone; that features distinctive performance styles; that blends oral and written transmission; that is performed as functional music, for celebration, and as concert music; and that lies at the intersection of multiple cultures and nationalities. It is a living, creative tradition that draws on oral history and archival material, bridging the trauma of pogroms, population displacement, the Holocaust, and Soviet repression. Practitioners, scholars, and students are hungry for the kinds of information and tools that MIR could provide.

6 CONCLUSIONS

In this paper, we have situated a community based music datafication project as an instance of 'non-commercial, publicly oriented initiatives aimed at enhancing human musical flourishing' (Born, 2020: 196). Furthermore, the development of, and ongoing praxis within KMDMP touches on all four of the dimensions of diversity identified by Born. The demographics of project leadership and participants is diverse with respect to gender, age, and nationality; the repertoire is outside of both popular and classical Eurogenic traditions; the foundational principles include the notion that music embeds social relations, memory, and history; and the real-world institution and community served are a nonprofit (the Klezmer Institute) and the community of klezmer musicians, dancers, teachers, students, aficionados, and scholars.

We would like to envision a world in which ISMIR and related initiatives support musical communities by making tools and resources available for cultural heritage documentation, preservation, and continuity in addition to supporting technology innovation. This would require a more cyclical engagement between the MIR community and music-makers in the communities that are their subject of study and analysis. Such a shift in focus would open further doors to addressing demographic diversity, make more varied repertoires available for documentation and study, and challenge MIR practitioners to grapple with ontological assumptions guided by cultural insiders.

The unique, difficult to datafy KMDMP corpus is a prime example of how community-based projects can create data sets for analyzing oral music as a 'total social fact' (Holzapfel et al., 2018). The new, post-revival klezmer musicians identified by project founder Anna Rogers as heirs to the music (see Section 3 above) are also important for documenting the heritage. Genre and performance metadata is implicit in the melodies for those who can perceive it, but is not explicitly written down precisely because the manuscripts were created by klezmorim who also carried this information as unwritten cultural knowledge.

Addressing Born's fourth dimension of diversity—engaging with real-world institutions and communities—has implications for both financing and collaboration. In typical crowdsourcing projects, volunteers do work that

benefits researchers. Both by necessity and by philosophy, KMDMP flips that model by foregrounding immediate benefits to musicians and the drive to build community over knowledge production or commercial interests. MIR-initiated funding streams for community-based projects that are (often) led by non-technologists (and which are likely to exist outside of academic and industry structures) would provide a critical form of support for cultural heritage projects, build communities eager to contribute to user experience testing for MIR tools, and feed datafied music back into the available analytical pool of musics of oral tradition. Though a non-profit organization, Klezmer Institute is not affiliated with any university or research center and therefore does not have access to funding that would typically be available to an academically affiliated MIR researcher or an MIR-focused research center. As far as we can see, no funding mechanism currently exists that would allow the Klezmer Institute to take advantage of the deep domain expertise within MIR or to allow MIR to learn from what we have to offer.

Datafication and technology are often perceived as an enemy of oral transmission, but the experience within KMDMP is quite the opposite. In fact, the KMDMP community has wholeheartedly committed to completing the digitization of the material and is eager to learn what computational tools can reveal about this repertoire. Collaboration with similar community and heritage-based projects has the potential to lead to technological development. Two potential avenues for exploration are creation of an open-source data ingestion pipeline for music scores similar to FromThePage (Brumfield and Brumfield, 2022) and a digital music (score) exhibition platform similar to the OMEKA project (Digital Scholar, 2022). Development of open-source ecosystems around such platforms has the potential to draw prospective students and specialists into the field by increasing the depth of engagement directly with diverse music making communities.

The potential for MIR itself to learn from engaging with the needs of these communities cannot be overstated. Constant user feedback—user experience research—is organic to any community-driven project, and centering the goals of documentation, preservation, and continuity within musical communities engages different imaginative muscles than those of observation and analysis. If music-making communities are valued as users of MIR tools, their feedback should be encouraged in the early stages of MIR software development. Holzapfel et al. (2018) present a simple MIR value chain that places MIR research and end-users at opposite ends. They say:

> A central problem is that only limited communication is established between MIR research and other parts of the value chain. This way, usually no feedback from users can be obtained regarding software that employs specific research ideas.[16]

Of necessity, the Klezmer Archive team and KMDMP leaders have become our community's MIR specialists. Software developers on the Klezmer Archive team are klezmer musicians themselves and, recognizing the vast and diverse community we wish to serve, the team has included a user experience expert from its inception.

Engaging directly with community-based projects like KMDMP is an exciting opportunity for the MIR community to find ways to support small projects with expansive goals that include research, performance, preservation, and other creative outputs. How can MIR support projects on a shoestring budget? Can MIR contribute to open-source ecosystems for crowdsourcing in music? Can MIR provide tools to support heritage groups to make their music accessible to both insiders and outsiders? Answering these questions by reaching out to diverse community spaces to ask their needs is the first step.

NOTES

1 Sholokhova (2006) and Deutsch (2016) provide history and context on Kiselgof and the An-sky expeditions. Selected ethnographic recordings with historical contexts are available in Volumes 2–5 and 10–11 of *Historical Collection of Jewish Musical Folklore 1912–1947* (Vernadsky National Library of Ukraine, 2001–2020).

2 Beregovski (2015, I47–I51) provides a biography of Makonovetsky.

3 Feldman (2016) and Rubin (2020) document histories and musical characteristics of klezmer music, including the history of the term itself. For ethnographies and essays on the contemporary klezmer scene, see Slobin (2000; 2002) and Walligorska (2013).

4 There are many projects that involve encoding historical music manuscripts in digital notation; see, for instance, the projects listed at the Music Encoding Initiative (MEI) website under "Community" and "Projects/Users" (Music Encoding Initiative, 2022). None of these, as far as we know, involve the kind of crowdsourcing and community engagement of KMDMP. The closest precedents for crowdsourcing music encoding are The Session (2022) and Traditional Tune Archive (2022). Both of these use ABC notation, which is not adequate for the complexity of the manuscripts in KMDMP.

5 Beregovski (2015) has 254 melodies and four concert pieces. There are additional and sometimes overlapping melodies in historical recordings: 42 commercial tracks by a group known as Belf's Romanian Orchestra (Fendler and Reid, 2007–2008); 466 tracks in ethnographic recordings from 1912–1947 (Vernadsky National Library of Ukraine, 2001–2020); and 614 tracks in ethnographic recordings from Moshe Beregovski and the Jewish ethnographer Sofia Magid (Grözinger et al., 2008). Correlating the notation and historical recordings will be part of the larger Klezmer Archive Project; see Section 5 below.

6 All authors contributed to the writing and research; section 3 is primarily by Christina Crowder and Clara Byom and section 5 is primarily by Yonatan Malin and Daniel Shanahan. Malin conducted the interviews found in the appendix. Crowder and Byom are the designers and administrators of KMDMP, while Malin and Shanahan are university-affiliated researchers in the fields of music theory and computational musicology.

7 Mishiro (2021) made a preliminary survey of this section of Beregovski's catalog for the first annual KMDMP digitizathon. Her overview indicates that of the 1,059 items listed, 834 are instrumentals and the remainder are vocal in nature: niggunim, song, and liturgical pieces. Her breakdown of instrumental titles by Feldman's (2016) classification system indicates 286 core; 57 transitional; 27 co-territorial; and 37 cosmopolitan pieces, with 427 pieces identified as march (35), tentsl (24), other (71), and untitled (297).

8 The observations here were made in consultation with project participants Hannah Ochner, Jutta Bogen, and Reuven Zaslavskyi.

9 Ilyefalvi (2018: 391–93) reviews the use of crowdsourcing in the digitization of folklore materials; see also Samiotis et al. (2020), and Ridge et al. (2021).

10 Ridge et al. (2021) articulate a similar ethic of mutual benefit, including the notion that through crowdsourcing, "communities find new opportunities to share their perspective and stories."

11 Notable progress is being made in this regard in the klezmer world, with several important festivals/workshops and a few influential musicians making gender equity in teaching spaces and on the bandstand a priority.

12 These four participants gave explicit consent for use of interview materials in publications.

13 As of October 2022, a collaborative team has begun the work of developing editorial policies for both music and text through the support of a National Endowment for the Humanities Scholarly Editions Planning Grant.

14 Tzanetakis et al. (2003) test the use of pitch histograms for genre identification, working at the interface between audio and symbolic MIR.

15 See, for instance, the key signatures in Berergovski (2015).

16 See further commentary in Born (2020: 194).

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendix.** Biographical Vignettes Of Four Project Participants. DOI: https://doi.org/10.5334/tismir.135.s1

COMPETING INTEREST

CC is Project Director and CB is Project Administrator for the Kiselgof-Makonovetsky Digital Manuscript Project (KMDMP). YM has been paid small fees (less than $100 per year) for his work on the Klezmer Archive Project Team, which is using the KMDMP as a testing ground. DS has no competing interests.

AUTHOR AFFILIATIONS

Yonatan Malin orcid.org/0000-0003-0684-8208
University of Colorado Boulder, US
Christina Crowder orcid.org/0000-0002-8771-8028
Klezmer Institute, US
Clara Byom orcid.org/0000-0003-4807-7528
Klezmer Institute, US
Daniel Shanahan orcid.org/0000-0002-6859-1324
Northwestern University, US

REFERENCES

Abe Schwartz Orchestra. (1920). *Alle in einem*. Recording, Columbia E7725. https://rsa.fau.edu/track/3517.

Beregovski, M. (1982). *Old Jewish Folk Music: The Collections and Writings of Moshe Beregovski* (M. Slobin, Ed.). Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9781512807516

Beregovski, M. (2015). *Jewish Instrumental Folk Music: The Collections and Writings of Moshe Beregovski* (M. Slobin, R. A. Rothstein, and M. Alpert, Eds.). Second edition, revised by Kurt Bjorling. Self-published, available from https://muziker.org/ (accessed July 27, 2022).

Bernard, C. A. (2005). *The Sound of Sacred Time: A Basic Music Theory Textbook to Teach the Jewish Prayer Modes*. Self-published, available from mailto:cantorbernard@gmail.com.

Born, G. (2020). Diversifying MIR: Knowledge and real-world challenges, and new interdisciplinary futures. *Transactions of the International Society for Music Information Retrieval*, *3*(1), 193–204. DOI: https://doi.org/10.5334/tismir.58

Brumfield, B., & **Brumfield, S.** (2022). *FromThePage*. Retrieved July 25, 2022, from https://fromthepage.com/.

Cohen, J. M. (2009). *The Making of a Reform Jewish Cantor: Musical Authority, Cultural Investment*. Indiana University Press.

CompMusic. (2022). *Computational Models for the Discovery of the World's Music*. Retrieved July 22, 2022, from https://compmusic.upf.edu/.

Deutsch, N. (2016). Thrice born; or, between two worlds. In J. Veidlinger (Ed.), *Going to the People: Jews and the Ethnographic Impulse* (pp. 27–44). Bloomington: Indiana University Press.

Digital Scholar. (2022). *Omeka*. Retrieved August 1, 2022, from https://omeka.org/.

Feldman, W. Z. (2016). *Klezmer: Music, History, and Memory*. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780190244514.001.0001

Fendler, A., & **Reid, R.** (2007–2008). *Transcriptions of All the Belf Tunes*. Retrieved July 26, 2022, from http://www.belfology.com/.

Frigyesi, J. (1993). Preliminary thoughts toward the study of music without clear beat: The example of 'flowing rhythm' in Jewish 'nusah.' *Asian Music*, *24*(2), 59–88. DOI: https://doi.org/10.2307/834467

Frigyesi, J. (2003). The unique character of Ashkenazi synagogal music: Homage to A. Z. Idelsohn. *Kenishta*, *2*, 147–166.

Gjerdingen, R., & **Bourne, J.** (2015). Schema theory as a construction grammar. *Music Theory Online*, *21*(2). DOI: https://doi.org/10.30535/mto.21.2.3

Gómez, E., Herrera, P., & **Gómez-Martin, F.** (2013). Computational ethnomusicology: Perspectives and challenges. *Journal of New Music Research*, *42*(2), 111–112. DOI: https://doi.org/10.1080/09298215.2013.818038

Grözinger, E., Hudak-Lazić, S., Schramm, M., & **Schirrmeister, S.** (2008). *"Unser Rebbe, unser Stalin …": Jiddische Lieder aus den St. Petersburger Sammlungen von Moishe Beregowski (1892–1961) und Sofia Magid (1892–1954)*. Wiesbaden: O. Harrassowitz.

Holzapfel, A., Sturm, B. L., & **Coeckelbergh, M.** (2018). Ethical dimensions of music information retrieval technology. *Transactions of the International Society for Music Information Retrieval*, *1*(1), 44–55. DOI: https://doi.org/10.5334/tismir.13

Horowitz, J. (1993). *The Klezmer Freygish Shteyger: Mode, Sub-mode and Modal Progression*. Unpublished manuscript. Retrieved November 7, 2022, from http://www.budowitz.com/Budowitz/Essays.html.

Huron, D. (1997). The Humdrum Toolkit: Reference Manual. Menlo Park, California: Center for Computer Assisted Research in the Humanities. Retrieved Jul 26, 2022, from https://www.humdrum.org/Humdrum/musedata2kern.html.

Ilyefalvi, E. (2018). The theoretical, methodological and technical issues of digital folklore databases and computational folkloristics. *Acta Ethnographica Hungarica*, *63*(1), 209–258. DOI: https://doi.org/10.1556/022.2018.63.1.11

Klezmer Institute. (2019). *Klezmer Archive Project*. Retrieved March 30, 2022, from https://klezmerinstitute.org/klezmerarchive/.

Klezmer Institute. (2022). *Kiselgof-Makonovetsky Digital Manuscript Project*. Retrieved July 22, 2022, from https://klezmerinstitute.org/kmdmp/.

Lipaev, I. (1904). Evreiskie orkestry. *Russkaia Muzykal'naia Gazeta*, 4, 101–103. Translated by Asia Fruman in *Articles, Translations, and Bibliographies* from The Klezmer Institute. Retrieved July 25, 2022, from https://klezmerinstitute.org/articles/.

Make. (2022). Retrieved August 3, 2022, from https://www.make.com/en.

Malin, Y., & **Shanahan, D.** (2022). *Mode and melody in klezmer music*. Seventh International Conference on Analytical Approaches to World Music.

Mishiro, M. (2021). *Jewish folk music of the early 20th century in Ukraine as seen in Zusman Kiselgof's manuscripts: Focusing on klezmer music*. KMDMP Anniversary Digitizathon (online). https://youtu.be/-0QSj9xdtbc.

Oudshoorn, N., Rommes, E., & **Stienstra, M.** (2004). Configuring the user as everybody: Gender and design cultures in information and communication technologies. *Science, Technology, and Human Values*, *29*(1), 30–63. DOI: https://doi.org/10.1177/0162243903259190

Powers, H., & **Wiering, F.** (2001). Mode, §I: The term. In *Oxford Music Online*. *Grove Music Online*. http://www.oxfordmusiconline.com/subscriber/article/grove/music/43718pg1. DOI: https://doi.org/10.1093/gmo/9781561592630.article.43718

Ridge, M., Blickhan, S., Ferriter, M., Mast, A., Brumfield, B., Wilkins, B., Cybulska, D., Burgher, D., Casey, J., Luther, K., Goldman, M. H., White, N., Willcox, P., Brumfield, S. C., Coleman, S. J., & **Prytz, Y. B.** (2021). Why work with crowdsourcing in cultural heritage? Chapter 3 in *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage – Community Review Version* (1st ed.). DOI: https://doi.org/10.21428/a5d7554f.1b80974b

Rubin, J. E. (2020). *New York Klezmer in the Early Twentieth Century: The Music of Naftule Brandwein and Dave Tarras*. Rochester: University of Rochester Press. DOI: https://doi.org/10.1017/9781787448315

Samiotis, I. P., Qiu, S., Mauri, A., Liem, C. C., Lofi, C., & **Bozzon, A.** (2020). Microtask crowdsourcing for music score transcriptions: An experiment with error detection. *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 901–907.

Schaffrath, H., & **Huron, D.** (1995). The Essen folksong collection in the humdrum kern format. Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Şentürk, S., Holzapfel, A., & **Serra, X.** (2014). Linking scores and audio recordings in makam music of Turkey. *Journal of New Music Research*, *43*(1), 34–52. DOI: https://doi.org/10.1080/09298215.2013.864681

Seroussi, E. (2009). Music: The "Jew" of Jewish studies. *Journal of the World Union of Jewish Studies*, *46*, 3–84.

Serra, X. (2011). A multicultural approach in music information research. *Proceedings of the 12th International Society for Music Information Retrieval Conference, pages* 151–56.

Serra, X. (2014). Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the 53rd AES International Conference on Semantic Audio*, pages 1–9.

Shanahan, D., & **Shanahan, E.** (2014). The Densmore Collection of Native American Songs: A new corpus for studies of effects of geography and social function in music. In *Proceedings of the 13th International Conference on Music Perception and Cognition*, pages 206–209.

Sholokhova, L. (2006). Jewish musical ethnography in the Russian empire: Ideology and chronology. In Nemtsov (Ed.), *Jüdische Kunstmusik im 20. Jahrhundert: Quellenlage, Entstehungsgeschichte, Stylanalysen* (Vol. 3, pages 227–232). Wiesbaden: Harrassovitz.

Slobin, M. (2000). *Fiddler on the Move: Exploring the Klezmer World*. New York: Oxford University Press.

Slobin, M. (Ed.). (2002). *American Klezmer: Its Roots and Offshoots*. Berkeley: University of California Press. DOI: https://doi.org/10.1525/9780520935655

Tarsi, B. (2017). At the intersection of music theory and ideology: A.Z. Idelsohn and the Ashkenazi prayer mode Magen Avot. *Journal of Musicological Research*, *36*(3), 208–233. DOI: https://doi.org/10.1080/01411896.2017.1340033

Tarsi, B. (2020). Uncovering the music theory of the Ashkenazi liturgical music: 'Adonai Malach' as a case study. *Analytical Approaches to World Music*, *8*(2). https://www.aawmjournal.com/articles/2020b/Tarsi_AAWM_Vol_8_2.html.

Tzanetakis, G. (2014). Computational ethnomusicology: A music information retrieval perspective. In *Music Technology Meets Philosophy: From Digital Echos to Virtual Ethos. Proceedings of the 2014 International Computer Music Conference*, pages 69–74.

Tzanetakis, G., Ermolinskyi, A., & **Cook, P.** (2003). Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research* 32 (2), 143–52. DOI: https://doi.org/10.1076/jnmr.32.2.143.16743

Van Kranenburg, P., de Bruin, M., & **Volk. A.** (2019). Documenting a song culture: The Dutch Song Database as a resource for musicological research. *International Journal on Digital Libraries* 20 (1), 13–23. DOI: https://doi.org/10.1007/s00799-017-0228-4

Van Kranenburg, P., Janssen, B., & **Volk, A.** (2016). The Meertens tune collections: The annotated corpus (mtc-ann) versions 1.1 and 2.0.1. *Meertens Online Reports*, *2016*(1).

Vernadsky National Library of Ukraine. (2001–2020). *Historical Collection of Jewish Musical Folklore 1912–1947*. Audio recordings. Retrieved July 26, 2022 from https://www.audio.ipri.kiev.ua/index.html.

Walligorska, M. (2013). *Klezmer's Afterlife: An Ethnography of the Jewish Music Revival in Poland and Germany*. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199995790.001.0001

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), Article 160018. DOI: https://doi.org/10.1038/sdata.2016.18

Zooniverse. (n.d.) *About*. Retrieved March 29, 2022, from https://www.zooniverse.org/about.

TO CITE THIS ARTICLE:
Malin, Y., Crowder, C., Byom, C., & Shanahan, D. (2022). Community Based Music Information Retrieval: A Case Study of Digitizing Historical Klezmer Manuscripts from Kyiv. *Transactions of the International Society for Music Information Retrieval*, 5(1), 208–221. DOI: https://doi.org/10.5334/tismir.135

Submitted: 04 April 2022 **Accepted:** 22 October 2022 **Published:** 30 November 2022