Final Report — White Paper

The Klezmer Archive Project

Project Director: Christina Crowder

Grantee Institution: Klezmer Institute, Inc.

Submission Date: April 30, 2023

NEH Digital Humanities Advancement Grant 2021-2022

Co-authors: Clara Byom, Eléonore Biezunski,
Christina Crowder, Dan Kunda Thagard,
Yonatan Malin, Max Rothman,
Matthew Stein, Schyler VerSteeg

THE KLEZMER INSTITUTE

Contents

This White Paper is an adaptation of the NEH DHAG Final Report delivered on March 31, 2023 formatted for a general audience.

Click Chapter and Appendix headings to jump to that page.

Acknowledgements

This white paper shows the work of the Klezmer Archive team over the course of a two year NEH Digital Humanities Advancement Grant between 2021-2022. In showcasing our work we would like to thank the many people who shared their time and expertise to help us along the way. First and foremost are the Advisory boards for both the Klezmer Archive Project (listed below) and the Klezmer Institute (rosza daniel lang-levitsky; Lisa Frankel; Emil Goldshmidt; Ethel Raim; Ilya Shneyveys; Mark Slobin). We have relied on advice from all of you to shape the project—tempering our tech ambitions with current best practice, and keeping our visioning of ontology and description aligned with the needs of the community.

Staff at the NEH Office of Digital Humanities (ODH) provided invaluable feedback on the draft Phase I application; as a new organization we were deeply grateful for the opportunity to connect with ODH professionals before we submitted our first proposal.

We would like to thank a number of individuals (among many consultants and advisors) who were particularly gracious with their time and who explicated key concepts early in the grant period. Jean Delahousse, one of the leads on the DoReMus project, led a tutorial session on the ontology and philosophy behind DoReMus that has had a lasting influence on the library team's approach. Daniel Shanahan was an early advocate for the project and helped us to get oriented in the current computational musicology and MIR landscape. Michel Biezunski of Infoloom has been incredibly generous, sharing not only his deep expertise in topic mapping, but making his ontology development tool Networker available to the team for a steeply reduced academic price.

Several librarians and archivists have been regular correspondents with the project. We'd like to thank Beth Dwoskin in particular for a number of thought provoking email exchanges. Judith Pinnolis, Beth Dwoskin, Amanda Siegel, and Jeanette Casey all shared their time for a series of wide-ranging conversations in preparation for a paper about how to keep librarians and archivists connected to users in digital spaces. We were fortunate to find Emese Ilyefalvi's 2018 survey of the digital humanities landscape for folklore archives early in the reseach phase. Her work has informed our approach to conceptualizing the MoOT ontology and the documentation of human relationships in the transmission of intangible cultural heritage, and her support has been of great assistance to the project.

This project would not exist without support from the klezmer community at large, and the KMDMP community in particular. There are too many to name who have touched the project, but we would like to call out some of the folks who were particularly helpful in

the first phase of the grant. Francis Harris, Grace van't Hof, Ilya Shneyveys, Joel Rubin, Jordan Hirsch, Sarah Schneider, Zilien Biret, and rosza daniel lang-levitsky all sat for lengthy "mental mapping" user experience interviews with Schyler VerSteeg and other team members. Hannah Ochner, Jordan Hirsch, Jutta Bogen, and Daniel Berkowitz were interviewed by Yonatan Malin for the TISMIR paper, sharing their experiences and insights about KMDMP.

We are also deeply grateful to the folks who come to our community meetings and programming around the project. Among the regulars, we'd like to thank Robert Blacksberg, Hannah Ochner, Ari Davidow, Sherry Mayrent, Uzi Smilovich, Daniel Carkner, Tobias Scheuer, Heike Scheuer, and Eric Rangell. Finally, we would like to extend an extra helping of gratitude to Sherry Mayrent for her enthusiastic encouragement from the beginning and unwavering faith in the importance of this project for the Jewish music community.

Klezmer Archive Advisory Board

NATIONAL ENDOWMENT FOR THE HUMANITIES

This project has been made possible in part by the National Endowment for the Humanities: Democracy demands wisdom.

1. Abstract

The Klezmer Archive (KA) project is creating a universally accessible digital archival tool for interaction, discovery, and research on available information about klezmer music and its network of contemporary and historical people. Taking individual melodies as the primary artifact, the digital archive will integrate existing tools and archival methods in novel ways to facilitate search and discovery rooted in the needs of its contemporary heritage community. Tooling and frameworks developed for the Klezmer Archive project will be available for heritage communities to adapt for their own domain-specific uses, and will be particularly useful for the preservation and study of intangible cultural heritage.

The project team is exploring newly-available open source knowledge engine technologies for organizing and validating heritage-based information, and is the first team to explore use of these tools in the digital humanities. The KA team is also the first (to our knowledge) to experiment with modeling music data directly in a knowledge graph.

The Klezmer Archive Project (KA) has completed a period of research and development supported by a two year Phase I NEH Digital Humanities Advancement Grant from 2021-2022. The primary goal of this phase was to produce a MVP (minimum viable product) plan identifying the core technical elements that must be integrated together to achieve the tool that could be called the Klezmer Archive as described in the project scope.

Outcomes from this phase include this white paper, a system architecture plan, a dev-blog, a landing page for the project, numerous public presentations, and a published journal article.

Quick Links

The best way to receive announcements and updates about the Klezmer Archive project is to sign up for the Klezmer Institute's (low-volume) Email Newsletter.

Project Page— klezmerinstitute.org/klezmerarchive/

Dev-blog — blog.klezmerarchive.org/

Code Repository — github.com/klezmer-archive

Public Meeting Recordings — 2021 Meeting and 2022 Meeting

KA Coffee Hours — The Archive as Network; Archiving MoOT

2. Project Summary

The Klezmer Archive (KA) project is creating a universally accessible digital archival tool for interaction, discovery, and research on available information about klezmer music and its network of contemporary and historical people. It will integrate existing tools in novel ways, and the project team is exploring newly-available open source tools for organizing and validating heritage-based information. This digital space will not hold physical objects, instead taking individual melodies as the primary artifact (as well as their expression through various media). Each melody artifact will be indexed using standard cataloging methods; described as a physical object using archive description frameworks (when applicable); surrounded by domain-specific metadata; rendered as machine-readable music notation formats; and, made available for user commentary, connection, and discussion.

The Klezmer Archive Project serves the humanities broadly by creating resources and tools for documenting, preserving, and researching the artistic expression of cultures and the people connected to them. Integrating musical search and analytical tools into a novel digital heritage resource will make a corpus of non-classical western music available to a wide audience of researchers in musicology, music theory, and ethnomusicology, with special relevance for computational analysis and corpus studies in the context of oral tradition. The tool will be available for heritage communities to adapt for use in their own domains and will be particularly useful for the preservation and study of intangible cultural heritage.

The Klezmer Archive Project has completed a period of research and development supported by a Phase I NEH Digital Humanities Advancement Grant in 2021-2022. The primary goal of this phase was to produce a MVP (minimum viable product) plan identifying the core technical elements that must be integrated together to achieve the tool that could be called the Klezmer Archive as described in the scope of the project. (Appendix A) The team created a blog and landing page that will be updated throughout all phases of the project. The team presented about the project in many forums on topics including: expanding ontologies for documenting musics of oral tradition; networks surrounding archives; developing innovative, community-based digital humanities projects on a small budget; and, the case study material (KMDMP) (Appendix J).

Bringing the Power of Knowledge Engines to Digital Humanities

One of the most exciting discoveries over the course of the grant period was the possibility to deploy the power of open-source knowledge engine technology together with domain-specific ontologies to give humanists more flexibility in the use of cutting-edge digital tools to assist in the collection, organization, and evaluation of cultural information.

Knowledge engines (also known as knowledge networks, knowledge graphs, reasoning platforms, reasoning engines, knowledge bases, or first-order logic systems) allow for the modeling of complex real-world data similar to how humans store knowledge—as a densely interconnected network of contextualized relationships.

The data model envisioned will allow researchers to store, view, and evaluate conflicting or mutually exclusive data in ways that are not well-supported through existing technologies; however, implementation of this advanced technology will not limit potential interoperability with existing collections and institutions. As the first team pursuing the application of open-source knowledge engine technology in the humanities, the KA project is committed to reducing barriers to entry for future projects. Tools like the Klezmer Archive will allow humanists, performers, and archivists to work together to tell the story of communities, arts, and cultures in a more comprehensive way than has been previously possible.

Test Corpus & Publishing Projects

Throughout the grant period, the team was involved with the ongoing digitization and ingestion of a case study corpus, informing the plan for transcribing and ingesting the entirety of the material as described in Appendix B. Additionally, a sub-group of the project team published a journal article, "Community Based Music Information Retrieval: A Case Study of Digitizing Historical Klezmer Manuscripts from Kyiv," in the Transactions for the International Society of Music Information Retrieval (TISMIR) about the first phases of digitization and ingestion of the case study corpus.

Project Team & Advisory Board

The team is composed of experienced software industry, archival, academic, and musical experts, most of whom are klezmer musicians themselves: Christina Crowder; Clara Byom, M.M.; Eléonore Biezunski, Ph.D.; Dan Kunda Thagard; Yonatan Malin, Ph.D.; Max Rothman; Matthew Stein; and Schyler VerSteeg. For the next phase of the project Andrew Parisi, Ph.D. will be joining the team. This project is advised by a team of renowned scholars, librarians, technologists, and cultural arts leaders: Jeanette Casey, M.M., M.L.S.; Jean Delahousse; Emese Ilyefalvi, Ph.D.; Mark Kligman, Ph.D.; Judith Pinnolis, M.M, M.S.; Pete Rushefsky; Amanda (Miryem-Khaye) Siegel; and Lyudmila Sholokhova, Ph.D.. David Lewis, researcher at University of Oxford e-Research Centre in Oxford and Lecturer in Computer Science at Goldsmiths, University of London, will be joining the advisory board for the next phase of the project. Throughout the Phase I grant period the team was regularly advised and supported by Michel Biezunski of Infoloom.

3. Origin

The UNESCO Convention in 2003 recognized the need to safeguard "intangible cultural heritage," including "practices, representations, expressions, knowledge, skills," and their surrounding artifacts. Preserving intangible cultural heritage such as music and folklore is inherently challenging because knowledge is shared via oral tradition and stored in the minds of performers and communities. Like others in heritage music interest communities, members of the klezmer community have dreamt of a centralized repository for klezmer tunes and their historical/ethnographic context for decades.

The project organizers were motivated to seek funding for this project after a series of impromptu conversations in December of 2019 as the team first formed. While a resource of this kind had been on the minds of many people independently, it wasn't until this specific group of people was assembled that it seemed possible to make it a reality.

4. Goals

Generating Useful Metadata for Practitioners and Researchers

As with other transnational and diaspora musics of oral tradition, documenting klezmer and its surrounding networks is complicated in distinctive ways:

- multiple alphabets and languages
- no singular composer
- inconsistent naming conventions
- multiple versions of tunes
- overlapping and sometimes contradictory genre classifications

Beyond accommodating the challenges listed above, to make the resource truly useful it will be necessary to:

- Open the "black box" of music to search "inside" musical items, such as by mode, characteristic rhythms or melodic figures, and other elements that can help musicians and researchers find tunes and identify tune families;
- Implement an ontology for organizing data specific to the domain and useful to performers such as dance form, social function, geographic range, and historically defined genre;
- Document relationships between people, especially in relation to the transmission of tunes and intangible cultural knowledge; and,
- Document valid, yet conflicting information.

Accommodating these needs is beyond the scope of traditional archival descriptive systems developed to document physical artifacts rather than abstract concepts like a musical melody and human networks. Our proposed solution is to use layers of metadata that will keep the artifact compatible with current catalog and archival systems, but will also expand to include metadata that make it comprehensible to practitioners and specialist researchers.

Viewing Musical Items as a "Total Social Fact" [1]

The Klezmer Archive project speaks to recent trends in the humanities that regard works of art and music not as self-standing aesthetic objects, but as entities that participate in networks of human relations. Through the active and ongoing work on the case study corpus, the team has observed the transmission of tunes through human networks in play-along sessions, workshops, presentations, and recordings. "Hannah's" Skotshne, which has been named after the most prolific digital notator and translator volunteering with the project, is one of several tunes that now circulate widely in the community. Through these observations of the case study corpus, the klezmer community broadly, and conversations in online spaces (such as Facebook Groups), key expansions of existing ontologies around the ideas of tune families and human relationships have emerged as highly valuable concepts for documenting intangible cultural heritage.

Serving Many Audiences

The Klezmer Archive will be an inclusive, welcoming space for anyone interested in the content regardless of experience. Anticipated users include scholars, independent researchers, teachers, musicians, composers, documentarians/film makers, dancers, community leaders, GLAM professionals, and individuals interested in eastern European Jewish culture. See Appendix K for an expanded list of anticipated users. While klezmer is the passion that unites the members of this project team, cultural practitioners and researchers throughout the world experience similar problems in capturing, organizing, and accessing data for music of oral tradition and folklore.

Interoperability & Extensibility

The Klezmer Archive project builds on the work of technologists, digital humanities experts, musicologists, folklorists, linguists, archivists, librarians, and others that have worked on related projects. This project is also a consequence of decades of research,

1. In a discussion of the ethics of MIR, the authors summarize work by Jean Molino and J. Nattiez, describing music as a 'total social fact:' "a human activity with sociological, historical, and physio-psychological dimensions, which has the potential to set in motion individuals, groups, and even the whole society." 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. https://doi.org/10.5334/tismir.13

preservation, collaboration, and creativity in the klezmer community. Projects in the humanities will be able to use and adapt all open-source tools and conceptual frameworks developed by the KA team, who will build upon numerous existing open-source tools and schemata (e.g. MEI, IIIF, TypeDB, etc.), and will contribute back to those ecosystems. Researchers working in the fields of folklore, music, language, ritualistics, folk dance, and any other field that documents the transmission of cultural knowledge will find tools and concepts useful for their projects.

5. Activities, Team, & Participants

The project team began 2021 by pursuing research in four interest areas in the DHAG work plan:

- Technology
- User Experience
- Fuzzy Tune/Genre Boundaries
- Editorial/Data Ingestion

Four overlapping work groups emerged:

- User Experience
- Library/Archive
- Corpus Studies/Musicology
- Tech

In addition to the independent investigation articulated for the working groups, the whole team engaged in a period of comparative analysis at the start of 2021. This allowed the team to gain further understanding of current work in the field; identify related projects; and, confirm our suspicion that no one else is pursuing a project like the Klezmer Archive.

User Experience Team

The team's User Experience expert, Schyler VerSteeg, facilitated team visioning exercises and led interviews with potential users. From an early visioning exercise a set of nine guiding principles for the Klezmer Archive MVP were established: open, access, outreach, analysis, reference, organization, standardization, metadata attribution, and curation. In fall 2021, the team engaged in a User Journey Visioning Exercise around potential users/use cases that further clarified the audience for the tool and their potential needs. (See Appendix E.) VerSteeg then conducted interviews with potential users to learn how they think about and relate to klezmer music, and what features they would like to see in the resource. A high-level report of the User Interviews is included in Appendix C. Schyler also facilitated meetings that helped the team align on a shared vision for the MVP.

Corpus Studies & Computational Musicology Team

The Corpus Studies/Computational Musicology Team led by Yonatan Malin held meetings to read articles, understand current academic research methods, and discuss concepts that would inform choices for the MVP. Team members Yonatan Malin, Christina Crowder, and Clara Byom collaborated to write an article about the Kiselgof-Makonovetsky Digital Manuscript Project (the KA case study corpus) that was published in Transactions of the International Society for Music Information Retrieval in November 2022. Computational musicologist Daniel Shanahan was a contributing author and also met with the whole team early in the grant period to discuss music encoding and analysis tools. This team was also engaged in giving feedback to the Library Team's ontology work and informed research regarding Music Information Retrieval for the Tech Team.

Library / Archive Team

The primary work of the Library Team (Christina Crowder, Clara Byom, Eléonore Biezunski) was to develop a preliminary ontology for use by the Klezmer Archive. The team researched ontologies used for organizing data in music library and archive systems and experimented in Infoloom's Networker to test concepts. Primary research questions included:

- How do we define "a tune"?
- How do we conceptualize tune families that group "variant of" melodies together using shared musical information?
- How do you seek out or search for a tune that you don't know, but that would pair well with another tune to make a medley?
- How do you document tune transmission in meaningful ways?
- How do you document and display competing narratives and differing opinions?
- How do we capture the conversations that happen in informal spaces, such as Facebook, lobbies at events, teaching spaces, and personal communications?
- How do you document meaningful cultural relationships between individuals that go beyond (or are not based on) particular items in the repertoire?
- How can we develop an ontology that serves our needs, while also interoperating with existing ontologies?

By the end of 2022, the team had made significant strides in developing an adaptation of DoReMus concepts for the new ontology, provisionally called Musics of Oral Tradition (MoOT). (Appendix H.) These ideas were presented in detail at the Music Library Association Conference (May 2022) and RILM Conference (April 2023). The library

team observed ongoing crowdsourced digitization of KMDMP materials, engaged with developing metadata vocabulary lists, recognized corpus-specific issues within musical notation and text, and identified participants to invite for further data processing.

Tech Team

The Tech Team (Max Rothman, Dan Kunda-Thagard, Matthew Stein) conducted research into prior art for data modeling and music encoding through extensive reading and conversations with outside experts. Team members held meetings with individuals working on related projects to learn about current work in the areas of generous interfaces, knowledge graphs, ontologies, models for validating different kinds of data, and the potential use of inference to derive richer information from the data sets. Primary research questions included:

- How might alternatives to RDF remove its limitations and open up more descriptive power?
- How do we make it user-friendly to query in complex domains (ex: uncertain dates)?
- How do we model and present conflicting information?
- How do we deal with levels of confidence, evidence, and building claims?
- How do we standardize this complex data?
- How do we model musical information (e.g. via MEI or MusicXML) and how do we relate that data to the rest of the knoledge graph?
- How do we allow researchers and users to follow citations from digitized data all the way to their primary source materials, including images or audio recordings?

Workflow

The team established regular monthly meetings, which allowed for individuals and working groups to report on their progress and for specific topics to be a focus of conversation. The team worked together to identify times for additional working group sessions, user interviews, and public meetings. Interteam communication took place in Slack for messaging and through Zoom video calls. Shared workspaces were created in Google Drive and Github. As with any distributed team, the challenge was to determine what conversations were best to have asynchronously and which required a dedicated video call to be most efficient. Three in-person working retreats were held during the grant period, which proved to be critical times of project development and team alignment. Working as a distributed team requires asynchronous communication and collaborative digital spaces.

6. Outcomes

This section covers the outcomes of the project in relation to its stated objectives. Subheadings will cover technology choices, interoperability, designing for FAIR & Open-source principles, and benefits to individual team members and the Klezmer Institute.

Final products of the DHAG award include:

- White Paper Report
- Architecture Plan for the software of the MVP (Appendix A)
- Plan for completing digitization and ingestion of the case study corpus (Appendix B)
- Summary Report of User Interviews (Appendix C)
- Project Dev Blog
- Journal Article for TISMIR
- Two community meetings/progress reports
- Conference papers & public presentations introducing the project to relevant fields.

Appendices D-I were developed for Phase II grant applications and are included here as further evidence of the research and development undertaken during the Phase I grant period.

Conference Papers and Public Presentations

Team members made presentations about the Klezmer Archive and its case study corpus throughout the grant period and beyond. Find the full list in Appendix J.

Project Dev Blog

The team launched a dev blog that includes the following blog posts as of spring 2023, and will be where the team continues to share the project's progress:

- Welcome to the Klezmer Archive Project Blog: An introduction to the project and its case study corpus, the Kiselgof-Makonovetsky Digital Manuscript Project
- The Fuzzy Tune Boundary Problem: A preliminary look at how we're thinking about the question "What is a tune?"
- Towards Documenting Human Relationships: First encounters with developing the "Human Relationship Concept"
- Documenting Folklore in Digital Structures — The Fuzzy Genre Boundary Conundrum: How do we make sense of conflicting, overlapping, or confusing genre

classifications?

PUBLIC MEETINGS

The team held public meetings to present our ideas and engage with questions from the community in June of each year. These can be viewed on the Klezmer Institute YouTube Channel: 2021 Meeting and 2022 Meeting. Two "KA Coffee Hour" presentations were held after the grant period to share conference papers with the general Klezmer Archive community: Archives as Networks presented a paper for the Music Librarians Association conference that discussed ontology issues; Archiving Music of Oral Tradition combined several papers and discussed issues around access to librarians, generous interfaces, and community connection. A third KA Coffee Hour session will be held in October 2023 to present the White Paper and System Architecture Plan.

Guiding Principles for the MVP/Architecture Plan

The following statements evolved over the course of the grant period. They inform technical decisions made for the architecture plan in Appendix A.

- This project is ambitious and risky due to its size, comprehensiveness, and complexity. To be sure the project can be completed, scope is trimmed when possible.
- The project's guiding principles are: openness, accessibility, outreach, analysis, reference, organization, standardization, metadata attribution, and curation.
- The team has a broad skill set and deep knowledge, so utilizing those strengths and adding team members where there are weaknesses is critical.
- A wide set of features must be available before the resource would be useful to its intended audiences. In early stages, it is critical that we use pre-existing tools as much as possible to cover that large feature area, while leaving room to add to or replace those tools afterwards when we have the bandwidth to add depth.
- We are not committing to any particular technology for the lifetime of the project, but have made preliminary technology choices to use in prototyping parts of the tool as the project moves forward.
- We will be intentional about building for the possibility of adopting other tools in the future.
- It is easier to ask new questions of a rich data model than to enrich a data model so that it can answer more interesting questions.
- We believe the open-world assumption will be useful for this project.
- Pre-existing ontologies and standards do not meet our needs alone, but our new ontology will be inspired by and/or adapted from DoReMus, RDA, the Music Ontology, CIDOC, FRBRoo, MARC, Library of Congress, and DACS.

Tool Choices & Building for Interoperability

The team values open source principles and sought out pre-existing tools where possible to solve its research questions and move the project forward quickly; to facilitate interoperability at various levels; and to allow the team to contribute back to other open source ecosystems. The team's research on data modeling led them to learn about Knowledge Engines through a series of consultations with logician and ontologist Andrew Parisi. The tech team developed and deployed a preliminary search tool for use on the KMDMP corpus in mid-2022. By the end of the grant period, the team decided to use the following pre-existing tools as a part of the architecture plan (see Appendix A):

- MEI for music notation data;
- mei-friend for last-mile music editing;
- IIIF Viewers and APIs for presenting primary source material images with basic metadata and annotations; and
- TypeDB for knowledge retrieval and reasoning.

Of these tools, Type DB will require the most thought in building for interoperability with existing popular data structures such as RDF. The project will evaluate user needs for RDF interoperability and build potentially multiple projections of the knowledge system data in RDF that are optimized for different user needs by making various tradeoffs in fidelity (see Appendix A; RDF).

In addition to work on the architecture plan, tech team member Max Rothman worked with Director Christina Crowder to develop data ingestion automations for the case study corpus, which will inform data ingestion pipelines for the KA tool.

FAIR & Open Source

Everything built throughout the entire lifecycle of the project—from software to content—will be freely available to technologists, musicians, and scholars alike, and anyone with the relevant knowledge will be encouraged to contribute. Current repositories (as of Spring 2023) are:

- klezmer-archive-blog: the code that runs the blog (public)
- rda-documentor: a tool for generating documentation for the RDA ontology (public)
- ontology: where the technical translation of the ontology will eventually live, currently just lots of work-in-progress notes (private)
- music-search: defunct music manuscript search prototype (private)

Contributors to code development will be credited in the GitHub repository and public announcements where applicable. Outside consultants will be identified as co-authors of text and/or code contributors. Score notations will be provided by volunteers under a contributor license agreement and licensed through a Creative Commons CC-BY 4.0 license with attribution to the holder of the source materials. Annotated digital images of physical sources will be shared via IIIF with a Rights Statement determined in collaboration with the source material holder. At later stages of the project, cooperative arrangements will be sought with archives and libraries with relevant holdings, and appropriate Rights Statements will be agreed with each cooperating institution.

Public Code developed by the team can be accessed at: https://github.com/klezmer-archive.

Outcomes for Team Members & The Klezmer Institute

This project has already benefited the team in many ways and we anticipate it will continue to do so in the future through the expansion of our network, publishing of articles, and skill-building for future work. Throughout the Phase I grant the team had numerous conversations with experts in related fields and projects, growing our network of colleagues substantially. This network will continue to be a source of support, ideas, and friendships for years to come for this project and beyond. As scholars and researchers (affiliated and independent), the publication in the Transactions for the International Society for Music information Retrieval by Malin, Crowder, Byom, and Shanahan benefited our individual careers. The tech team also sees this as a direct benefit to their careers in the future, as this is an opportunity for them to learn from working on a large system with unusual needs and to become experts on applying knowledge system technology in humanities fields, qualifying them to be consultants for similar projects in the future.

The Phase I Digital Humanities Advancement grant also benefited the Klezmer Institute because it was the first grant application the organization submitted. Securing this funding boosted the Institute's prestige and has helped us to secure other grant funding, including an NEH Scholarly Editions Planning Grant for the Kiselgof-Makonovetsky Digital Manuscript Project.

7. Evaluation and Impact

The Klezmer Archive Project succeeded in meeting its research and discovery goals in this phase and the primary objective of creating an architecture plan for the MVP (Minimum Viable Product) for the Klezmer Archive tool was achieved.

The project work plan established a timeline to conduct research in related fields; to learn about the current state and limitations of a variety of technologies in digital humanities; and, to attract a team of knowledgeable consultants and advisors to support the core project team in various ways. The team was successful in this endeavor, developing an extensive comparative analysis rubric, a Zotero library, and a bibliography of related projects and key researchers/project leaders. Some of the most important (among many) related projects are the Music Encoding Initiative (MEI), mei-friend, International Image Interoperability Framework (IIIF), and DoReMus (Doing Reusable Music).

Because the DHAG I was envisioned as a planning grant, much of the impact of the project will be felt in the years to come and can only be anticipated now. To our knowledge, knowledge engine/first order logic systems have rarely—if ever—been used in the humanities, which for the most part relies on RDF and related technologies. The KA team may be one of the first teams with the technical expertise and commitment to implement knowledge engines for humanities work. The team will also be the first (to our knowledge) to model music directly within a knowledge graph.

Interim Reporting and Feedback

Two community meetings were held as intermediate reporting points that facilitated feedback from the user community. Team members consulted with members of the KI and KA advisory boards during the grant period, and sought out colleagues in the core technical disciplines that emerged during the research and discovery phase. In 2023, three public "KA Coffee Hour" sessions were (will be) held on the following topics:

A. Developing an ontology for Music of Oral Tradition (MoOT)
B. Using digital humanities tools to connect the archive to community
C. Presenting the KA architecture plan and white paper

User experience interviews proved to be a useful opportunity for feedback and helped ensure the project's continued alignment with community needs.

Community Engagement

Public outreach strategies included user experience interviews, mental mapping interviews, conference presentations, librarian and archivist interviews, conversations with specialists in related fields, blog posts, and "community meetings" (reporting on project progress and inviting questions from attendees). Community engagement is an inherent component of work on the case study materials because KMDMP uses a crowdsourcing model. The Ashkenazic expressive culture community is eager to see this project come to fruition and happy to contribute in any way invited. The project team included a user experience expert from the start, so testing and user engagement were planned into the project from the beginning, and plans were made intentionally flexible to accommodate research results.

Peer Networking

The team also benefited greatly from conversations with a number of specialists in our research areas.

- Yonatan Malin's colleague **Daniel Shanahan** kindly shared several early conversations with the team on MIR and computational musicology topics that identified key projects and technologies. Shanahan is a co-author on the TISMIR paper.
- **Michel Biezunski** of Infoloom gave the team access to his topic mapping platform "Networker" at a deeply discounted academic rate and consulted extensively with the library team on topic-mapping philosophy and ontology development. Biezunski remains a close friend and advisor to the project.
- **Jean Delahousse**, one of the founders of the DoReMus project led the team through an in-depth tutorial on the DoReMus ontology that had a profound impact on a library team that was more accustomed to MARC and archive descriptive systems such as DACS. Delahousse will join the Project Advisory Board for Phase II.
- **Andrew Hankinson** and **David Lewis** generously shared their time in fall 2022 to share the current status of MEI and mei-friend and talk through planned developments in those tools. David Lewis will join the Project Advisory Board for Phase II.
- Logician and ontologist **Andrew Parisi** (an industry colleague of Max Rothman) introduced the team to knowledge engine technology and led the team to discover a related open source tool—TypeDB. Parisi will join the project team for Phase II.

Project Strengths

The Klezmer Archive is a project of the Klezmer Institute (KI)—a digital-first nonprofit 501(c)(3) organization founded in 2018 to support Ashkenazic expressive culture through research, teaching, publishing, and programming. The Klezmer Institute's position as a newcomer, outside of typical academic structures and GLAM spaces is both a strength and a challenge. Because of financial constraints, the Klezmer Institute has adapted low-cost/

low-code business platforms to automate workflows and as an interim data store. While not reliable for long-term FAIR compliance, these tools can be of great utility in the intial design phases for community-based projects working with limited budgets. Likewise, without legacy systems to accommodate the team has been able to rely on its strengths—industry experience, academic depth, domain expertise, and community connections—to do a lot of blue-sky thinking.

Even as our project builds on and learns from existing projects and tools, the team's research revealed that the project is unique, ambitious, and cutting-edge on multiple fronts:

- The size of this project is monumental and no project of similar scale aims to combine musical search with both expert and user-contributed metadata and commentary.
- This project will be the first (to our knowledge) to model music data directly into a knowledge graph.
- To our knowledge, this will be the first project in the digital humanities space that will use inference/reasoning.
- Musics of Oral Tradition will be the first ontology that highlights the transmission and transformation of intangible cultural heritage as it is shared through communities.

The team understood from the start that the project is extremely ambitious in scope. However, eight out of nine core team members are themselves klezmer musicians/scholars who understand the core problems the project is seeking to address and are deeply invested in seeing this project to fruition. Further, the case study corpus (KMDMP) is a community-driven digital humanities project that has involved over 200 volunteers. This community surrounding KMDMP has been an integral part of the conversations that are guiding development of the Klezmer Archive and they are eager to contribute further. As a result of this combination of circumstances, we have been able to think expansively and develop a plan for a resource that has the potential to fundamentally impact the way archives, libraries, research institutions, and communities of scholar-practitioners document, preserve, and research the artistic expression of cultures and the people connected to them.

Project Challenges

The KA team encountered challenges with regards to time, communication, and resources, but in every instance members learned and adapted to meet the needs of the project. For each team member, working on this project has been an add-on to paid work and other creative projects that take considerable time, so being intentional about setting aside appropriate time for independent and group work on the Klezmer Archive was critical to the team's productivity. The project operates on a limited budget, which factored into choices

made for the planning and future implementation of the project, as well as immediate investments in personnel, administration/management tools, travel, and conference/program attendance. A larger budget during this grant period would have allowed for compensation of more of the team's work (they graciously donated hundreds of hours above and beyond the grant support) and facilitated more in-person meetings for small working groups and the whole team.

Defining the Audience for Digital Humanities Tools in Intangible Culture Heritage

The core audience for the project was refined and expanded during the grant period. The impact of the resource on klezmer and Ashkenazic expressive culture practitioners, researchers, and community leaders was clear from the start, as having a centralized repository for klezmer tunes and their historical/ethnographic context has long been a dream of the community. The team also developed an expanded list of professionals, students, and other communities that would benefit from access to an organized resource around klezmer music (Appendix K).

The potential benefits for the documentation of intangible cultural heritage has become more clear over the course of the grant period. In addition to its technical goals, the Klezmer Archive seeks to become a central hub for its community of musicians, audience, researchers, and educators in the same way that resources like The Session, Hymnary, and the Country Dance and Song Society have become the connective tissue of their respective communities. Looking forward, the project team will seek meetings with representatives from these projects to inform the community-facing design process for the Klezmer Archive.

In addition to innovations around the ontology for Musics of Oral Tradition and incorporation of knowledge engines, the work of combining existing technologies (e.g. for MIR, image display, archival description) should not be overlooked. Working in these shared ecosystems is sure to enhance the Klezmer Archive Project and contribute to the advancement of other projects in the humanities broadly. A tool that successfully integrates these existing tools with knowledge engine technology, corpus-specific ontologies, modeling music in the graph, user commentary, and cross-tagging on items will become a model for other projects seeking to document heritage communities and folklore.

8. Next Steps: Visioning Phase II — Prototyping & Testing

The Klezmer Archive project team is dedicated to seeing this project to fruition. The team has continued to work on the project after the conclusion of the DHAG Phase I grant while simultaneously pursuing additional funding opportunities. The primary research questions the project continues to engage with are:

- How do we model creative, folkloric, and historical traditions data in useful and meaningful ways for practitioners and researchers?
- How can user interfaces displaying knowledge graph information be built that support the needs of both the most expert and the most novice users?
- How can interfaces be built that allow for simultaneously searching using multiple media types? (e.g. text, knowledge graphs, musical notation, musical audio, etc.)
- How can user interfaces be built that support serendipitous discovery (browsing) alongside multifaceted search?
- How do we continue to engage the community in building and creating an archive?

During the next project phase, the team will prototype and test tools and concepts through an iterative process involving community feedback, consultation with specialists, and experimentation. The team anticipates developing and testing feature prototypes based on an initial prioritization:

- P1- Music search, data modeling/inference;
- P2- Score ingestion, metadata ingestion, metadata search;
- P3- Artifact view/presenting scores.

Ultimately, prototype success will be determined by rigorous testing and iteration. To evaluate the project's progress and success overall, the team will invite domain experts (e.g. archivists/librarians, musicians, musicologists, historians) and the Advisory Board for probing, high-level conversations. As tools are shared with users, the team will seek feedback via conversations, surveys, and UX research.

The upcoming project phase will include intense periods of editing/cleaning of the case study data for ingestion. After the grant period, the team expects to launch an initial version of the archive tool to the public with the entirety of the KMDMP manuscripts incorporated. Later stages of the project will address questions of how to engage institutions with klezmer music collections, and how to incorporate and process digital material shared with the Klezmer Archive project.

We encourage other researchers to be in touch as they grapple with similar questions. We are happy to consider collaborations, explorations, and experiments with others as our work continues to move forward.

Creative Impact

Beyond its impact on research, the Klezmer Archive will impact the artistic output in the community, as has been observed in the circulation of the KMDMP corpus. Numerous musical projects have emerged since the start of KMDMP, including tune folios, new musical collaborations, weekly play-along sessions, and recording projects. While it is true that the manuscripts released in KMDMP included over 1,300 newly accessible tunes, the Klezmer Archive serves the community in a similar way—revealing materials that are otherwise disparate and difficult to access. We anticipate that this resource will inspire artists around the world to engage with this essential cultural legacy in new and exciting ways.

Guide to Appendices:

Appendix A — Architecture Explanation / MVP Plan

This architecture plan identifies key components of the Klezmer Archive resource and how they will work together as currently envisioned based upon the case study corpus, the Kiselgof-Makonovetsky Digital Manuscript Project. The gray items on the diagram represent features that are part of the overall plan and will be further defined in later project phases. As of Spring 2023, our next long-term goal is to get the central components below functioning together expediently, and allow for changes in the future to scale up or adapt to other needs.

The tool will consist primarily of a knowledge system surrounded by a constellation of knowledge system writers and knowledge system readers.

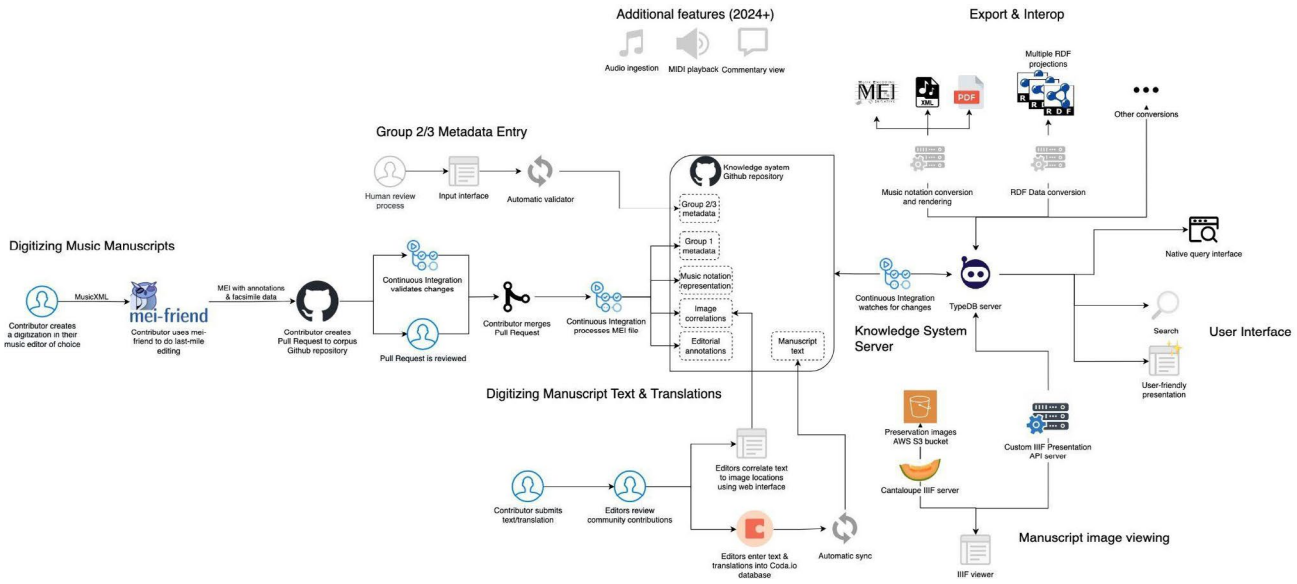

Figure 1. Architecture plan for the Klezmer Archive MVP.
Find a full-scale image at https://tinyurl.com/2sdr5xvd

1. The Knowledge System

Canonical Data Storage

The source of truth for data in the knowledge system will be plain text files containing a textual knowledge notation stored in a Github repository. Files in the repository will not typically be edited by people, but instead written by various programs that act as knowledge system writers. This approach allows us to use Git to track and audit the history of the data as it evolves over time. It also provides extremely strong durability guarantees: it's practically impossible that our data could be lost through a database going down, getting corrupted, etc. This approach will not scale to a high rate of writes (hundreds per second), but we predict that our write rate will peak at 1 per second, and on average be much lower. If

the project grows to the point where our write rate exceeds Github's practical limits, using a textual notation for the data makes it easy to write programs that can migrate the data to another storage medium.

This data storage will contain all the data we intend to model in this phase of the project, which is described below.

2. Knowledge System Data

The subsections here describe individual items on the System Architecture Plan.

Music Notation Representation

Music notation will be stored as a graph representation of MEI XML. Semantically XML is a tree, and graphs are a strict superset of trees, so with this approach we will be able to use the full range of MEI's representational power while also bringing the knowledge system's full capabilities to bear in representing metadata, including allowing us to connect data directly to parts of notation, make automatic inferences about notation and its connected data, and allowing for queries that include both metadata and notation parameters.

The translation between MEI XML and this graph representation can be performed automatically, allowing us to easily round trip notation through the knowledge system and make use of MEI's strong ecosystem of tools, like Verovio and mei-friend.

Manuscript Text

Text appearing in the manuscripts, translations of that text, and context about the above will be stored in the knowledge system. Together with musical notation data, this data forms the primary source material of the tool. We are considering multiple approaches for modeling text, including TEI.

Image Correlations

Links between musical notation data and text to the manuscript images on which they appear will be stored in the knowledge system in a format compatible with the IIIF Presentation API. This data allows users to drill into manuscript digitizations to see the images those digitizations were made from. In the future, we intend to digitize further types of non-musical manuscript documents such as catalog pages, and correlations between those digitizations and the source images will use the same facility.

Editorial Annotations & Metadata

Annotations added by editors in the process of cleaning manuscript digitizations will be

stored in the knowledge system so that the context they add can be understood by users viewing visualizations of the information. For example, an editor could disambiguate a note that is unclear in a handwritten score and enhance that disambiguation with the musicological reasoning supporting their decision.

Metadata will be added by project editors in four main categories (Appendix G).

Group 1 Metadata

Group 1 includes all metadata recorded in traditional catalogs (often using MARC, DublinCore, etc), standardized subject area categories, and any other factual data that relates to the item. This includes labels such as title, composer, informant, collector, source, date, etc., and can also include archival descriptive information (e.g. of a physical manuscript or gramophone disc using DACS).

Group 1 metadata will include metadata about transcribed manuscripts. This data will be modeled using our ontology and stored in the knowledge system, connected to the notation it references.

Group 2 and 3 Metadata

Group 2 metadata refer directly to the artifact but include classifications that rely on both observable facts (e.g. time signature, cadential formula, title, marginalia) and subjective classifications that a curator/editor adds to the data to articulate groupings, typologies, and genre distinctions that are specific to the cultural understanding of a musical item (or an item in general). Examples include dance form, social/ritual function, and musicological characteristics which are not easily (usefully) captured in simple keyword fields. Additional useful Group 2 metadata can be inferred from outside sources, for example, instrumentation and personnel on historical recordings through reference to discographies and other scholarly research.

Group 3 semantic metadata allows items to be linked to each other by documenting relationships, and for those relationships themselves to become items (nodes). The project team is expanding and adapting schema and ontologies developed in the RDF space to describe music of oral tradition, and to capture human relationships that are essential to understanding cultural transmission.

Group 2 and 3 metadata includes information that is produced through a higher level of interpretation by human experts. Once we have developed an ontology for modeling this information, we will create a process for ingesting this data at a later stage of the project.

Knowledge System Querying

For information in the knowledge system to be presented to users, it needs to be stored in a queryable format. When new changes to the system are added to the Github repository, a Continuous Integration (CI) job hosted in a 3rd party CI service (e.g. Github Actions, CircleCI) will translate the changes into the appropriate set of commands to insert the data into a persistent TypeDB server. (See Appendix I for a comparison of RDF, Property Graphs, and Knowledge Engines, such as TypeDB.) Data in this server can then be queried efficiently by various knowledge system readers to power user-facing experiences.

Manuscript Image Viewing

To make manuscript images available in an accessible and performant manner, the project will adopt the IIIF Image and Presentation APIs.

IIIF Image Server: Cantaloupe

Cantaloupe is a high-performance dynamic image server that implements the IIIF Image API. The project team will deploy a Cantaloupe server that serves the manuscript images. The Image API is concerned only with the details of images, their sizes, qualities, etc. and thus requires no data from the knowledge system to function.

IIIF Presentation Server

The Presentation API deals with the relationships between images, the context in which they are to be understood, and other information that is desirable to have modeled in the knowledge system. A knowledge system reader will be developed that implements the IIIF Presentation API.

IIIF Viewer

The IIIF community has already developed a variety of viewers that consume the Image and Presentation APIs. The project will choose one or more viewers that meet our needs and deploy them on a public website.

User Interface

The project will develop one or more user interfaces that present the data in the knowledge system in a user-accessible manner. These interfaces will allow users to browse, search, and explore the project's data, powered by a suite of knowledge system reader servers developed by the project.

Search

The tool will allow users to search for all of the data stored in the knowledge system,

including manuscript text, music notation, metadata of all groups, and any combination of those. The natural methods of searching those different types of data are different enough from each other that the project will likely need to develop a tool dedicated to performing those searches dynamically and performantly.

Browse

In contrast to search, the presentation of the data for browsing is likely to be relatively static. In Phase II, the team will explore building interfaces to display this data guided by user research and the principles of generous interfaces.

Digitizing Musical Manuscripts

To ensure that the digitizations included in the knowledge system are high-quality, they will go through a tool-assisted but human-driven editing/cleaning process.

Volunteer-Contributed Digitizations

To ensure the widest reach for our pool of contributors, we have and will continue to allow them to contribute digitizations in whatever music editor they prefer (e.g. MuseScore, Sibelius, etc.). Standardizing on a single music editor would add friction to potential contributions, and building a custom editor is out-of-scope for this stage of the project. There are a number of tradeoffs with this approach:

All popular music notation editors support exporting in MusicXML, not MEI. As mentioned above, we plan to store music notation data in an MEI-compatible format. The style of MusicXML exported varies widely between editors. We will need to standardize the data for us to be able to create consistent user experiences.

Compared to MEI, MusicXML has limited or no capability to model metadata, editorial markup, unreadable or ambiguous notation, or correlations between musical data and manuscript images. We will need a way to collect that data from a source outside of MusicXML. These challenges are addressed in the later stages in the digitization pipeline.

Digitization Editing/Cleaning

The project will recruit and train an editorial team who will be responsible for cleaning and standardizing digitizations. The majority, if not all, of the editorial team will be members of the Klezmer Archive Team or KMDMP Scholarly Editions Team. The team will use the process and tools detailed below.

MEI-FRIEND

mei-friend is an open-source "last-mile" MEI editor which allows users to convert MusicXML to MEI, alter the imported MEI XML, and enhance the MEI XML with additional MEI-specific features like metadata, editorial markup, image correlations, etc. The editorial team will take the MusicXML produced by volunteers and import it into mei-friend, where they will enhance it with those MEI-specific features. The result of this process will be a "packed MEI XML" file that contains all metadata about the digitization that will be modeled in the knowledge system, including data that will be modeled using other non-MEI schemas (e.g. image correlations, which might be better modeled using the Web Annotations schema). mei-friend is a general MEI tool and not specific to our project: it can only produce MEI files, and we believe it will be more expedient to work within that limitation than to fork it or build a new tool from scratch. Thus MEI will be used to "carry" that information from mei-friend to the next step, where it will be extracted. We believe that under that approach, there will be significant overlap between the needs of our project and the needs of the MEI community at large, so we plan on making open-source contributions to mei-friend to close gaps in its current functionality that are needed for our project.

Github Pull Request

Once a member of the editorial team has taken a first pass at cleaning a digitization, they will make a Pull Request to a Github repo where cleaned digitizations are stored with the MEI XML file produced by mei-friend. That Pull Request will be reviewed by other members of the editorial team, assisted by various automated tools running as CI jobs:

- Jobs that perform checks to ensure standardization criteria are met, that metadata is well-formed, etc.
- Jobs that assist reviewers by presenting data in a more digestible format, for example rendering MEI XML as musical notation in the Github Pull Request web interface

Once all the checks pass and the Pull Request has been approved by sufficient reviewers, it is merged.

Insertion Into The Knowledge System

When a pull request to the digitization repo is merged, CI jobs will run which extract the data from that "packed MEI XML" file and insert it into the appropriate places in the knowledge system Github repository. The original MEI XML file will also be retained in the digitization Github repository for auditing purposes. Since by this point the metadata in the MEI file has been standardized, automated extraction and transformation of that data should be straightforward. Once inserted into the knowledge system, the digitization will be available to the other systems that read from it.

Digitizing Manuscript Text And Translations

Similar to digitized musical manuscripts, volunteers contribute digitizations of manuscript text and their translations to the project, and those contributions are reviewed and cleaned by an editorial team that will be recruited by the project. Unlike with musical notation, the data involved with text and translations is just simple plain text data, so the toolchain for managing it can also be simpler. It requires two parts: a database interface accessible to nontechnical users, and an interface that can be used to correlate text with the manuscript images on which it appears.

Coda.io Database

coda.io is a web-based document tool that combines features of databases, spreadsheets, and no/low-code app builders. For the purposes of this project, it will be used as a nontechnical-user-friendly database tool where the editorial team can input, review, and discuss volunteer-contributed text and translations. The database features coda.io offers allow the editorial team to work in the schema of the knowledge system without having to understand its technical details.

Once text/translations have been approved by the editorial team, a custom-built integration will sync the data into the knowledge system Github repository.

We intend to replace Coda.io with a purpose-built tool once the schema and ingestion processes are better-developed.

Image Correlation Interface

The editorial team will need a way to create, edit, and review correlations between digitized text and the location of that text in manuscripts. The solution must meet the needs of the editorial team: technically powerful and flexible, but friendly to nontechnical users. Planning for this tool is still in progress, but the team has considered various avenues: many IIIF viewers have Web Annotation viewing capabilities and some even have authoring capabilities. Another approach might be to reuse mei-friend to create MEI XML files that contain only text and no musical notation data.

Group 2/3 Metadata Entry

The project will build a process for ingesting Group 2 and 3 metadata during a later stage of the project, building on the experience gained from working with the music manuscript and text data, and through work in Coda.io. The process will likely involve a similar mix of human and automated checks and user-friendly interfaces.

Export and Interop

While the user interfaces for the project will meet the needs of nontechnical users who want to interact with the data in the knowledge system, expert users like academics, libraries, archives, and programmers will want to be able to interact with the raw data contained in the knowledge system. The project will make that data available on-demand in a variety of formats so as best to meet the needs of those users.

Native query interface

APIs for querying the knowledge system in its native data format will be made available, as will tools allowing technical users to easily make use of those APIs.

Music Notation

The music notation stored in the knowledge system will be semantically equivalent to MEI XML but not actually XML itself. The project will implement a simple web-based tool that converts the data from the knowledge system's format into plain MEI XML, making it available to users on-request. Similar tools will be built for other formats: PDFs of music notation rendered via Verovio, MusicXML through a similar tool to the MEI one previously mentioned, etc. The project will evaluate user needs for other formats as well and will build or utilize existing converters to meet those needs.

RDF

The data model used by the knowledge system is strictly more expressive than that of RDF, thus providing RDF interoperability with the data will necessitate some loss of fidelity, either by complicating the RDF schema with reification or by not including all the data available in the knowledge system. The project will evaluate user needs for RDF interoperability and build potentially multiple projections of the knowledge system data in RDF that are each optimized for different user needs by making different tradeoffs on this fidelity problem. These projections can be performed in bulk and made available as a single download or can be performed on-demand and made available through the web.

Other Conversions

The project will evaluate other user needs for interoperability and consider building other conversions or projections of the knowledge system data informed by those needs.

Additional Features

Several desired features are not in-scope for the current plan, such as: including audio data, MIDI playback of music notation data, capturing expert commentary on the data (Group 4 metadata). We believe that by focusing on the aforementioned features first will build a strong foundation on which features like these can be built in the future.

Appendix B — Case Study Digitization & Ingestion Plan

This document articulates the work of individuals & communities involved in ingesting data and the Make ingestion pipeline for the Klezmer Archive Phase I Case Study Corpus. Parts of this process are additionally articulated in this article: *Community Based Music Information Retrieval: A Case Study of Digitizing Historical Klezmer Manuscripts from Kyiv* co-authored by Yonatan Malin, Christina Crowder, Clara Byom, and Dan Shanahan.

About The Case Study Corpus

The Kiselgof-Makonovetsky Digital Manuscript Project connects musicians, Yiddishists, and scholars today with klezmer musicians from more than 100 years ago through an international community-driven digital humanities project. To date, the project has engaged more than 200 registered volunteer participants and has reached many hundred workshop participants and audience members. As of spring 2023, 1,150 of 1,300 tunes have been digitized, nearly all text has been transcribed/transliterated, and these once long-lost pieces are now being performed around the world.

About Music Data in the Corpus

The music manuscripts found in the Kiselgof material include a small number of transcriptions from audio sources (wax cylinders), notebooks written by individual musicians (Motl Reyder, Gershkovitz, and others), and loose sheets of handwritten music notations assembled into folios —author(s) unknown, but ascribed to klezmer musicians. The Makonovetsky manuscript was written by a single author—a klezmer violinist—encompassing approximately 230 pages of music with extensive text passages and marginalia that include observations about the social function of the music and performance practice.

About Text Data in the Corpus

The musical corpus contains a wide variety of textual information including marginalia added by the original author, annotations by multiple later catalogers on the manuscripts, song lyrics, liturgical and paraliturgical texts. These texts are mostly written in pre-YIVO Yiddish, pre-revolutionary Russian, and Hebrew in Ashkenazic dialect, along with small amounts in other languages such as German and Italian.

Workflow Scenarios

The subsections that follow articulate different workflow scenarios using a standard format.

Scenario 1. Volunteer Registration

Who: Volunteers who wish to join the KMDMP commons are registered to the project, which gives them permission to access the Google Drive folders and documents that coordinate music and text work on the corpus.

Goal: Streamline manual processes to set permissions, to get buy-in to community standards, and (in the future) to sign a release form for contributed scores and other material. To have an accurate list of project members.

Process: Make.com automation process designed by Max Rothman and Christina Crowder is triggered by a Google form completed by a prospective volounteer:

1. Volunteers register for the project using a Google Form
2. The make.com Participant Registration Scenario onboards new members
    a. Sends introductory email
    b. Adds members to KMDMP Google contact list
    c. Sets permissions on Google sheets and folders
    d. Adds participants to email lists based on their choices
    e. Adds participants to a Google group
    f. Flags members who would like an orientation

Scenario 2. Volunteer-Contributed Music Data

Who: Volunteers with KMDMP create initial digitizations of the music found in images of handwritten manuscripts.

Goal: Create an initial data set for later editing. Return immediate musical outputs to volunteer contributors and project members. Provide resources that can be used to develop community around the performance and teaching of this musical corpus.

Process Part I:

1. Volunteers are invited to contribute digital renderings of music in the handwritten manuscripts.
2. Volunteers use the score notation software of their choice to digitally notate the music.
3. Volunteers are asked to reproduce what is on the page and to add notes about unclear/obscured notes and other editorial decisions that are made during the digitization process.
4. Volunteers are asked to save the score in three formats: .pdf, .native, uncompressed MusicXML

5. Volunteers upload scores via a Google Form which triggers the Make.com Score Processing Scenario automation process.

Process Part II:
The Make.com automation scenario designed by Max Rothman and Christina Crowder accomplishes the following tasks :

1. Standardizes file names by setting variables entered in the Google intake form including KMDMP number, tune title, notator name.
2. Routes files to various google drive folders by file type
3. Attaches updated versions and sets version to stay indefinitely to preserve all iterations
4. Sends confirmation email to uploader and project administrator
5. Sends error notifications to a dedicated Slack channel.

Scenario 3: Music Data Standardization

Who: An small editorial/ingestion team drawn from the Klezmer Archive and KMDMP Scholarly Editions editorial policy teams.

Goal: To transform volunteer-contributed scores into standardized, annotated, machine-readable renderings of the original manuscripts, and to add appropriate KA level 1 metadata to the MEI files using the open-source "last mile" editor mei-friend.

Process:

1. Uncompressed music .xml files are imported into mei-friend.
2. The editor will add/edit Level 1 metadata in the MEI file, including KMDMP number, title, digital engraver, MEI editor, etc.
3. The editor will verify the imported xml against the manuscript to correct any errors in the initial digitization and make corrections as necessary.
4. The editor will annotate musical information, such as unclear/obscured notes, repeat bars in the score, unusual music symbols, indefinite repeat and segno sections, later annotations on the manuscript, and any other irregularities in consultation with both the original manuscript and the digital renderings (and any included notes).
5. The cleaned MEI will be validated by automatic checks to ensure consistency with our standards.
6. The cleaned MEI for each item is verified visually by additional editors (e.g. visually through a Verovio.org .svg rendering).
7. Approved MEI rendering is inserted in the graph and input files are archived.

Scenario 4: Volunteer-Contributed Text Data

Who: Volunteers with KMDMP create initial digitizations of the text found in images of handwritten manuscripts.

Goal: The community was asked to digitize all the text, translate, and transliterate to the best of their abilities. This work was often done collaboratively, bringing several experts together. When encountering unclear, obscured, or difficult to decipher items, digitizers were asked to note this in the document.

Process:

- Volunteers are invited to contribute translations and transliterations of text in various languages and alphabets.
- Google sheets were created to mimic the presentation of the manuscript and catalog images: 01 Beregovski Catalog Tracking Sheet; 02 Heft/Mak Tracking sheet.
- Volunteers enter data in cells that correspond to elements that are found on the manuscript pages organized by catalog and KMDMP number (title, notes, other observations, etc.).
- A third sheet was created so that non-Yiddish and Russian speakers could request translation assistance on text to be included in digital scores. (04 Translation Wish List).

Scenario 5: Text Data Standardization

Who: Members of the KMDMP Scholarly Editions editorial team

Goal: To edit volunteer digitized text to create an standardized, authoritative set of text data as represented on the manuscript pages.

Process:

1. Develop an expanded data storage system in coda.io to enable editorial work on text and metadata. This system facilitates cross-linked data, workflow management, complex search and filtering, creation of data sub-sets, sophisticated data views, and web-presentation options.
2. Ingest existing Google sheets into the new coda.io system.
3. Develop a workflow tracking system for text elements organized by KMDMP number to facilitate translation and transliteration standardization, addition of notes and commentary as appropriate, and an approval process for various text elements.
4. Text editors proceed with standardization of text data.
5. The approved standardized text data will be inserted into the Knowledge System and input files will be archived.

Scenario 6: Correlating Music and Text Data to Manuscript Images

Who: KMDMP Scholarly Editions Editorial Team and Klezmer Archive Team Members

Goal: Define regions on the images that correlate to musical or textual data.

Process:

1. Adopt the IIIF Image API via the Cantaloupe image server, making all KMDMP manuscript images available at permanent URLs.
2. Adopt the IIIF Presentation API using provisional handwritten data and make the manuscript images accessible and navigable via an IIIF viewer. This allows us to move forward with this process before the knowledge system is ready to be integrated with.
3. Implement tools allowing editors to correlate image regions with musical text and data – this happens during the prototyping phase.
4. Editors tag regions and correlate to either music or text data as held in the data store.
5. Implement a knowledge system reader that replaces the provisional handwritten data used for the IIIF Presentation API with data derived dynamically from the knowledge system, which will be its source of truth.

Scenario 7: Adding Additional (Group 2) Metadata

Who: Core members of KMDMP community, SE and Klezmer Archive teams.

Goal: To create a basic set of Group II metadata (genre and typology) for the corpus.

Process:

1. Add metadata tags in coda.io to items that pertain to genre and typology including (but not limited to) social function, dance genre, key, mode, etc.
2. The approved additional metadata will be inserted into the Knowledge System and exports of the original data from coda.io will be archived.

Appendix C — Mental Model Interviews- High-Level Report

Nine in-depth interviews were conducted with community members during the grant period. Interview participants presented a range of experience levels, instruments, and backgrounds, as well as in their focus and interests within Klezmer. These interviews allow the Klezmer Archive team an understanding of user mental models (i.e. the ways in which community members think about and interact with Klezmer), as well as the ability to incorporate input from a wide variety of users to inform both the scope and form of the resource.

While further analysis and application from these interviews is forthcoming, the following high-level insights have emerged:

- A constant across virtually all participants is the resonance of the cultural and historical importance of klezmer, and the draw to this music being at least in part an emotional one. Both the listening experience and the connection to history shape this relationship for community members. This is expected to help shape both to how the resource can be utilized at the feature level, and also the way it should "feel": how content is written and organized, branding and language, how the community is engaged in the growth of this resource.
- Interviews highlighted potential applications for both teachers and students of Klezmer. As future ideation and design takes place we should consider use cases for both of these audiences, and maintain the understanding that many users may be both teachers and students in different contexts.
- Three of the foundational attributes of the Klezmer Archive project mission are openness, access, and outreach - making previously inaccessible information and resources available to all, and serve as an effective place to enter the Klezmer community. Our interviews reflect these principles by highlighting the need for the Klezmer Archive to account for the experiences of novices, in this case serving as a "front door" to klezmer. In practical terms, we must explore whether this necessitates different platforms for novice community members and experts, or the creation of one resource approachable by both.

Appendix D — Bibliography

Anttonen, Pertti. "Lost in Intersemiotic Translation? The Problem of Context in Folk Narratives in the Archive," n.d., 18.

Bakka, Egil. "Revisiting Typology and Classification in the Era of Digital Humanities." ARV Nordic Yearbook of Folklore 75 (2019): 153–79.

Blumauer, Andreas, and Helmut Nagy. *The Knowledge Graph Cookbook*, n.d.

Born, Georgina. "On Musical Mediation: Ontology, Technology and Creativity." *Twentieth-Century Music 2*, no. 1 (March 2005): 7–36. https://doi.org/10.1017/S147857220500023X.

Born, Georgina, and Andrew Barry. "Music, Mediation Theories and Actor-Network Theory." *Contemporary Music Review*, 37:5-6 (2018): 443-487. https://doi.org/10.1080/07494467.2018.1578107.

Bue, Edited Margrethe Støkken, and Annika Rockenberger. "Notated Music in the Digital Sphere," n.d., 168.

Cuthbert, Michael Scott, and Christopher Ariza. "Music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data," 2010, 6.

Gell, Alfred. *Art and Agency: An Anthropological Theory*, (New York: Oxford University Press, 1998).

Hogan, Aidan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, José Emilio Labra Gayo, et al. "Knowledge Graphs." ArXiv:2003.02320 [Cs], January 23, 2021. http://arxiv.org/abs/2003.02320.

Ilyefalvi, Emese. "The Theoretical, Methodological and Technical Issues of Digital Folklore Databases and Computational Folkloristics." *Acta Ethnographica Hungarica* 63, no. 1 (June 2018): 209–58. https://doi.org/10.1556/022.2018.63.1.11.

Kranenburg, Peter Van, Jörg Garbers, Anja Volk, Frans Wiering, Louis Grijp, and Remco C. Veltkamp. "Towards Integration of Music Information. Retrieval and Folk Song Research." Department of Information and Computing Sciences, Utrecht University Technical Report UU-CS-2007-016 (2007). http://webdoc.sub.gwdg.de/ebook/serien/ah/UU-CS/2007-016.pdf.

Malin, Yonatan, Christina Crowder, Clara Byom, and Daniel Shanahan. "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*, no. 1 (November 30, 2022): 208–21. https://doi.org/10.5334/tismir.135.

Martin-Rodilla, Patricia, Martín Pereira-Fariña, and Cesar Gonzalez-Perez. "Qualifying and Quantifying Uncertainty in Digital Humanities: A Fuzzy-Logic Approach." In *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, 788–94. TEEM'19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3362789.3362833.

Nisheva-Pavlova, Maria, and Pavel Pavlov. "Ontology-Based Search and Document Retrieval in a Digital Library with Folk Songs1." Edited by Ana Baptista, Peter Linde, Niklas Lavesson, and Miguel A. Brito. *Information Services & Use* 31, no. 3–4 (October 16, 2012): 157–66. https://doi.org/10.3233/ISU-2012-0645.

Pivert, Olivier, Olfa Slama, and Virginie Thion. "Fuzzy Quantified Queries to Fuzzy RDF Databases." In 2017 *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–7. Naples, Italy: IEEE, 2017. https://doi.org/10.1109/FUZZ-IEEE.2017.8015632.

Saitanov, Ilya, Evgenia Khazdan, and Sofia Vlasova. "Klezmer Music Modes: Two Approaches to Corpus Markup." In *Language, Music and Gesture: Informational Crossroads*, 2021.

Salaba, Athena, Tanja Merčun & Trond Aalberg. "Complexity of Work Families and Entity-Based Visualization Displays" In *Cataloging & Classification Quarterly*, 56:7 (2018) 628-652.

Sharpe, K. (2004). "Alfred Gell, Art and Agency: An Anthropological Theory." European Journal of Archaeology, 7(2) (2004): 209-213. https://doi.org/10.1179/eja.2004.7.2.209.

Shaw, Ryan. "Modeling and Reasoning about Incomplete, Uncertain, and Approximate Historical Dates." 2021. https://uncch.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c5c6333e-0f4f-4cd0-9fbe-ae2f002a35b5.

"Time Ontology in OWL," n.d. https://www.w3.org/TR/2020/CR-owl-time-20200326/.

VanHandel, Leigh, and David Temperley. "Introduction to the Second Special Issue on Corpus Methods." *Music Perception* 31, no. 3 (February 1, 2014): 191–191. https://doi.org/10.1525/mp.2014.31.3.191.

"Vaticle TypeDB (Formerly Grakn): Schema Concepts." Accessed September 5, 2021. https://docs.vaticle.com/docs/schema/concepts.

Verhoeven, Deb, and Toby Burrows. "Building the Australian Knowledge Graph: HuNI (Humanities Networked Infrastructure)," n.d., 8.

Webster, Gemma, Hai Nguyen, David E. Beel, Chris Mellish, Claire D. Wallace, and Jeff Pan. "CURIOS: Connecting Community Heritage through Linked Data." In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 639–48. Vancouver BC Canada: ACM, 2015. https://doi.org/10.1145/2675133.2675247.

Wiering, Dr F. "WITCHCRAFT Project Final Report." Final Report, 2010.

Wiering, Frans, Tim Crawford, and David Lewis. "Creating an XML Vocabulary for Encoding Lute Music," n.d., 8.

Weissenberger, Lynnsey K. "Linked Data in Music and Its Potential for ITMA/Traditional Music Web Resources." *Brio* 55, no. 1 (2018): 52–57. https://doi.org/10.5281/ZENODO.1002056.

Appendix E — User Journeys

The following are examples from a User Journey Exercise conducted at our October 2021 retreat, which explored potential ways the Klezmer Archive tool could fill current spaces in resources/tools available to individuals in the klezmer community, and allow the Klezmer Archive team to prioritize goals and features based on this exercise. These are potential features for the team to consider.

User Journey #1: Musician in a gigging band

The musician is trying to put together a concert program, but wants to find some new tunes to create medleys with tunes they already know. They are able to search extensively using domain-specific keywords, regions, and musical qualities to find tunes that pair well. These tunes already have sheet music that can be added to a setlist and shared with bandmates. Since the tunes are linked in the graph, there is contextual information and commentary already available to the musician.

User Journey #2: Workshop Teacher

Teachers in the klezmer community can create their own workspaces with playlists, charts, and additional resources to share with their classes. Each tune they choose will have charts that can be transposed into any key for any instrument. Since the tunes are linked in the graph, contextual information and commentary are available to the teacher and student.

User Journey #3: Community Band Member

An individual decides they want to play klezmer, but don't know where to start. When they begin searching online they come across the Klezmer Archive. There they find a directory with a list of community bands, camps, workshops, and events. Some of the bands and events have playlists created by the leaders that are linked to the tunes with the Archive. After perusing the directory and associated playlists, the individual decides to contact the closest community band leader and ask how to join!

User Journey #4: World Music Student

The student has two weeks to write a paper about klezmer music. When they search the internet, the Klezmer Archive comes up and they find a collection of pages that answer fundamental questions (e.g. what is klezmer music, etc.). A curated list of articles, books, and top scholars is provided to give students direction for continued research. Automated citations are provided.

Appendix F — Datafication Challenges in the KMDMP Corpus

Figure 2 shows a sample from the KMDMP case study materials that demonstrates the challenges of digitally encoding these handwritten manuscripts.

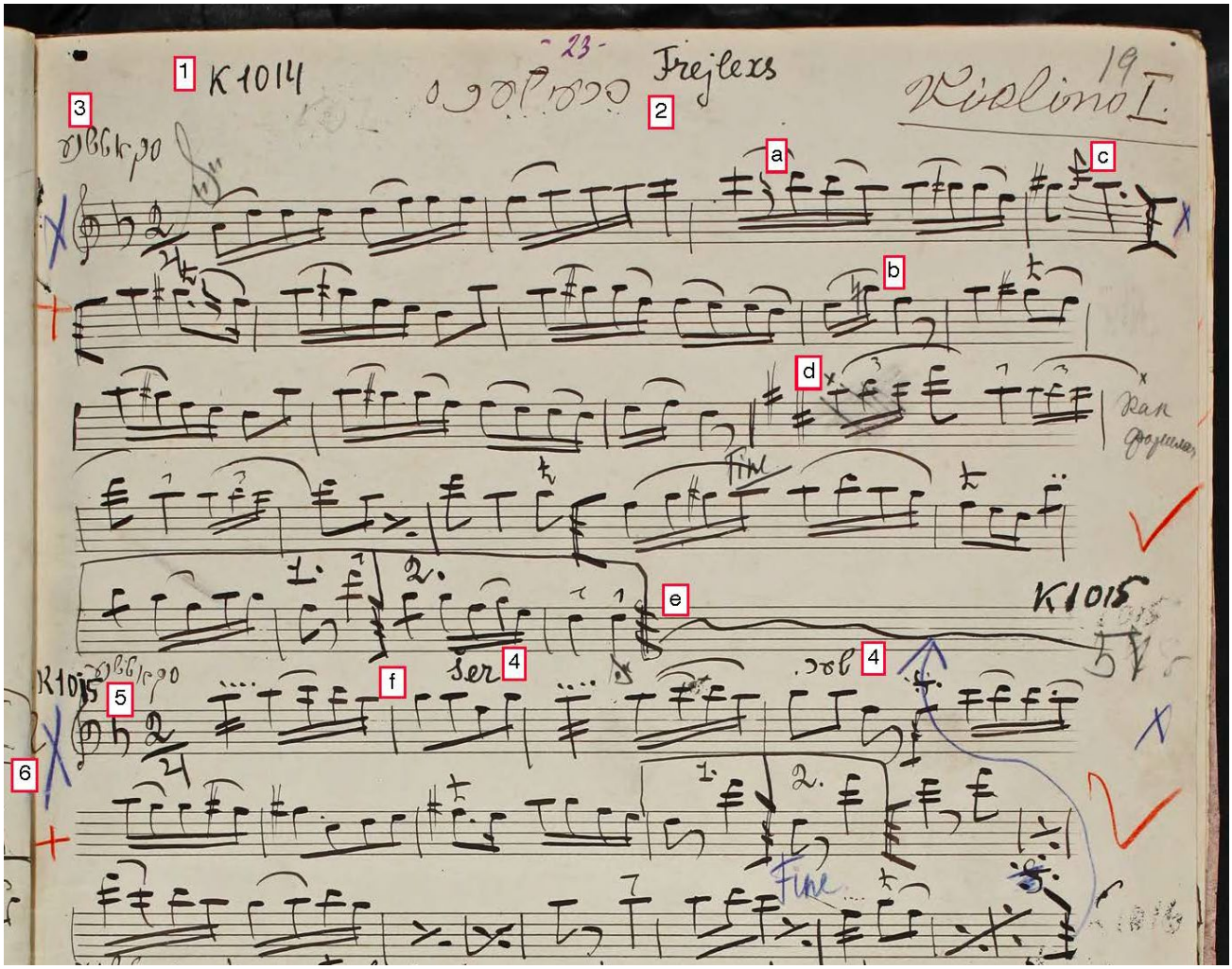

Figure 2. Annotated image of KMDMP 02-37-1014, and 02-37-1015. Numbers show language issues; letters show musical issues.

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

1. 'K1014' shows Beregovski's catalog number. The second tune below is marked 'K1015'.
2. The tune is identified by genre as a '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 script, in

different ink and another hand.

3. The tune is also identified in Yiddish script as a skotshne. This may be the designation made by the informant, 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 Roman and Yiddish script as a sher, in what appears to be Beregovski's handwriting.
5. It is also identified in Yiddish script 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. The likely pitch for a krechts 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 an ordinary Segno.

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>.

Appendix G — Four Metadata Layers

Layer 1: Traditional Catalog Information

Layer one includes all metadata recorded in traditional catalogs (often using MARC, DublinCore, etc), standardized subject area categories, and any other factual data that relates to the item. This includes labels such as title, composer, informant, collector, source, date, etc. Due to the transnational and orally transmitted qualities of klezmer, documentation of specific klezmer tunes is complicated in several ways that are shared with other musics of oral tradition: a) multiple alphabets and languages; b) no singular composer, inconsistent naming conventions, and multiple versions of tunes; and c) unknown dates. These issues will be present in all metadata layers and have known solutions.

Layer 2: Domain-Specific Metadata

Layer 2 metadata refer directly to the artifact but include classifications that rely on both observable facts (e.g. time signature, cadential formula, title, marginalia) and subjective classifications that a curator/editor/culture bearer adds to the data to articulate groupings, typologies, and genre distinctions that are specific to the cultural understanding of a musical item (or an item in general).

Development of robust domain-specific metadata allows users to search for melodies in ways that are not dependent on title or composer/artist. Examples include dance form, social/ritual function, and musicological characteristics, which are not easily (usefully) captured in simple keyword fields because a typology descriptor is essential to contextualizing words and descriptors that have specific cultural meaning—particularly when the same word can be used in more than one classification (e.g. “freylekhs” as both dance form and rhythmic texture), when genre boundaries are “fuzzy,” and when genres and typologies can overlap.

Additional useful metadata can be inferred from outside sources, for example, instrumentation and personnel on historical recordings through reference to discographies and other scholarly research. Development of an open ontology framework for domain-specific metadata is inspired by recent work by Egil Bakka and the HuNI project (See Appendix D). Phase II work will include experimentation with typologies and genre schema that will facilitate this kind of search and will usefully organize what appear to be conflicting information about melodies. The team will also investigate how to incorporate structures that will allow users to create novel typologies to suit their own purposes.

Domain-Specific Metadata Case Study
The item in Figure 3, Mak 1-32-117, indicates multiple, potentially conflicting classifications supported by musical information in the manuscript, context in the text marginalia, and klezmer-specific domain expertise.

Title: Gasn shtikl; Zhokl

Performance genre: Gasn nign; Dobranotsch (non-dance genres)

Performance style/texture: dobriden/dobranotsch; waltz; zhok (mutually exclusive)

Social Function: Gasn nign (on the street/public); Dobranotsch (escorting guests)

Knowledge of the klezmer idiom indicates that "tempo waltz" does not mean that the piece should be played as a waltz, and that use of the word *zhokl* indicates a meaning more like the Romanian *joc* (dance tune) than the klezmer texture *zhok*.

Further examples of classifications include: mode(s) in use, tetrachord(s) in use, modulation, cadential formula, rhythmic pattern, schema, harmonic basis (modal vs. harmonic). Typology examples include: Feldman Category (core, transitional, co-territorial, cosmopolitan); Beregovski Category (music for dancing, music for listening).

Layer 3: Semantic Metadata

Semantic metadata allows items to be linked to each other by documenting relationships, and for those relationships themselves to become items (nodes). The project team is expanding and adapting schema and ontologies developed in the RDF space to describe music of oral tradition, and to capture human relationships that are essential to understanding cultural transmission (see Appendix H). Key inspiration for ontology development to date have been DoReMus and HuNi (see Appendix D), but the team is exploring current research in folklore, dance studies, and related fields. As with user-initiated typologies described in layer 2, the project team is interested in facilitating user-generated relationships, e.g. "tune x works well in a suite with tune y," "Dave mentioned tune x on a Facebook thread about modular tunes."

Layer 4: Commentary & Discussion

The Commentary and Discussion layer captures information about artifacts that is not usefully recorded in semantic language. At present, this kind of knowledge is shared in community spaces such as Yiddish culture festivals, workshops, and Facebook groups, or through scholarly publications. Though critically important for the continued creative and scholarly development of klezmer, none of these spaces generate data that is both organized and searchable, or that attaches specifically to the music under discussion. Commentary may relate to one or more items, and may relate and/or connect any other metadata. There

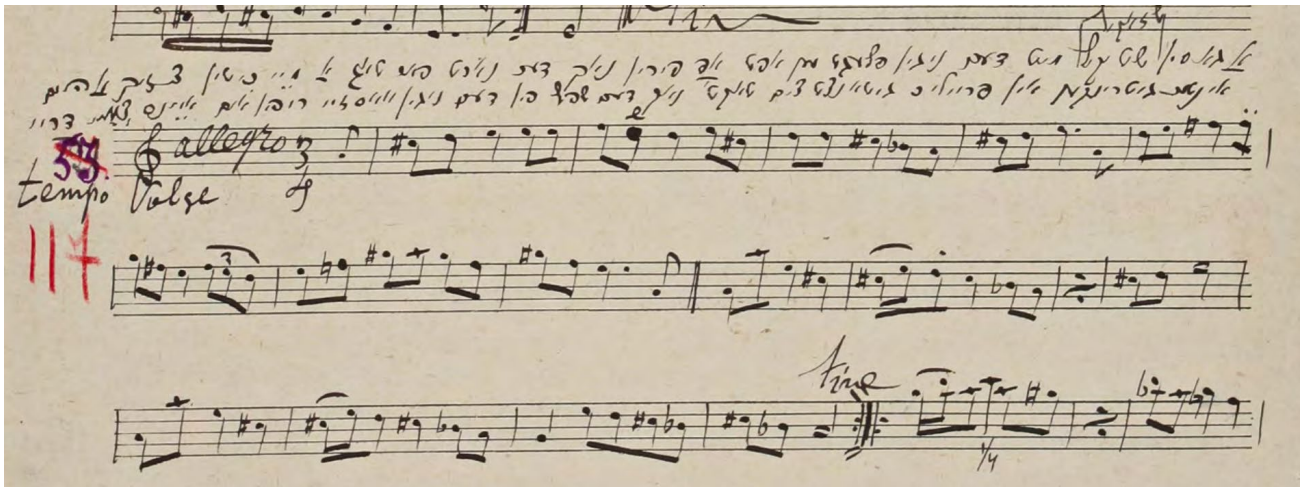

Figure 3. KMDMP Mak 1-32-116, Gasn shtikl (zhokl)

are many models for capturing comments and even making comment text searchable from blogs to entire social media platforms. For digital humanities tools, organizing comments and commentary around individual items is a way to surface discussion (and disagreement) between scholars in a public space, to invite practitioners to share stories and experiences about tunes and the people they learned them from, and to encourage newcomers to engage with the material in an embedded context. While such engagement cannot replace direct person-to-person transmission of culture, a digital repository of user-generated cultural and scholarly knowledge attached to the data store of musical items has the potential to dramatically expand access to and engagement with both the music and the living culture that supports it.

An overarching concept: The Talmudic View

An early vision for how to create a generous interface to display metadata and commentary on an item in what the team calls the "Talmudic" View. In book form, the page layout for this ancient Jewish commentary on the Torah (old testament) centers a core subject text, which is surrounded with commentaries by different sages and includes location data and references to related text/commentary in the margins. In a similar way, the KA team envisions a musical item (ms, digital notation, or sound clip) surrounded by commentary, with those given the most weight by the editorial/curatorial process/algorithm easily visible in the primary page view, and others on subsequent pages. Metadata from "layers" 1-3 would be available in a header/footer/sidebar.

Allowing scholars and practitioners to contribute both factual data and more subjective classifications and commentary will create a space for equally valid, but sometimes competing narratives that more closely represents the way that culture bearers understand this music. While this layer of metadata is vital, it realistically will not be included until a later stage of the project and will be informed by targeted UX research.

Appendix H — Preliminary Expansions for the Musics of Oral Tradition (MoOT) Ontology

Through experiments within Infoloom's Networker and study of the DoReMus ontology, the team has found two potential points of expansion for the KA ontology (tentatively called Musics of Oral Tradition or MoOT): tune concepts and human relationship concepts.

The Tune Concept & Tune Families

Within the framework of the DoReMus ontology (itself adapted from CIDOC and FRBRoo), a "tune concept" is an arbitrary but identifying label that describes a melody (usually a performance or instance of a melody) that can have many expressions that directly relate to that performance. For example, a single performance released on a gramophone record could have many different "expressions" such as a reprint 78 with a different 'B' side; re-issue on Cassette or CD; a YouTube video; sheet music in a tune folio. All of these expressions can be connected to a single performance via the tune concept. A historical recording of the "same" tune by a different artist would be articulated as a different "tune concept" but would be tightly linked to the other through a simple "variant of" relationship. Many variants of a core melody associated through "variant of" relationships can be identified as a "tune family."

Deciding what precisely constitutes an identifiable melody, or "a tune," is a question the team has identified as the "Fuzzy Tune Boundary" problem. This question can be illustrated through an experiment to capture information in a Facebook post from November 2021 as a way to apply the DoReMus "triad" of Work/Concept -Expression-Event to documenting real-world conversations about traditional music. The initial post posed a question about tunes that share parts with other tunes, or "modular" tunes. In the thread, 13 commenters discussed 8 different tune concepts identified as potential "variants of" each other to varying degrees. Sub-threads identified two other distinct examples of "modular" melodies. The melody that instigated the conversation (Chasem Senem) happens to have particularly transnational variants, but it is hardly unusual. In this case variants appeared in Turkish, Greek, Crimean, Klezmer and Sephardic music, and as instrumental, song, opera, and symphonic expressions.

Understanding connections between tunes is a topic of great interest in the klezmer community but also with those interested in many other traditional musics. The ability to surface melodic similarity, variants, and shared parts through both computational analysis of the music, and through semantic metadata will allow researchers and performers to understand cultural transmission and exchange through time and across geography in a way

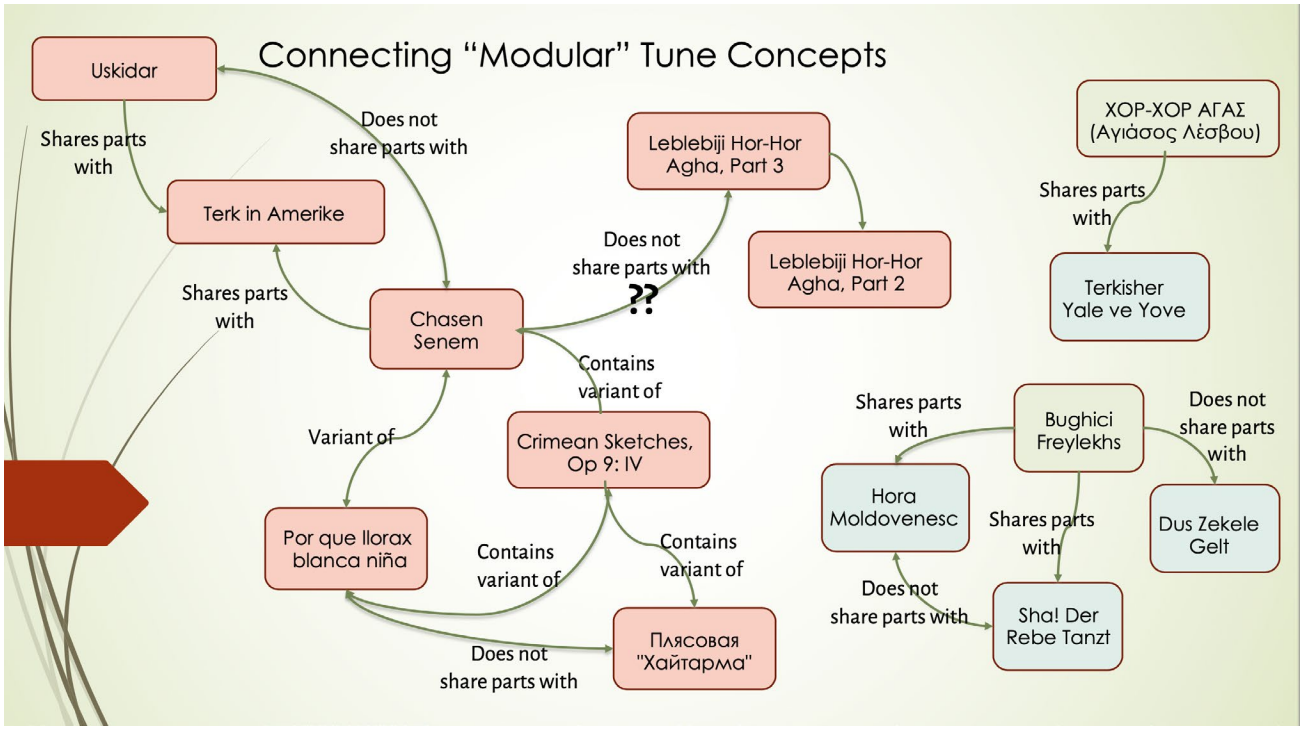

Figure 4. Connecting "Modular" Tune Concepts, prepared for MLA Conference, 2022

that has not been previously possible. The goal is not necessarily to find a "solution" to the fuzzy tune boundary problem by imposing crisp (potentially artificial) distinctions between tune artifacts, but rather to gain a deeper understanding of how music creation responds to the culture it is embedded in and to trace the networks of human actors that transmit this culture.

Cultural/Human Relationship Concept *[2]

What is currently being referred to as the "human relationship" concept allows people to be connected in a way that more accurately reflects how cultural knowledge is exchanged—through mentorships, within families, between bandmates, via private lessons.

For example, we set out to document field recordings from the Michael Alpert Field Recording Collection and related items. When looking at just a single song, we can see that Michael recorded Bronya Sakina singing the song on a cassette in March of 1986 and we know that Michael's ensemble recorded it on an album years later. When we connect the two versions via a song concept, we understand that they are related, and a liner note tells us that Michael learned the song in Bronya's kitchen. While this is accurate, it does not fully represent the quality of the human relationship. To reflect this, we can connect Michael

2 The final term is TBD, but in the section below it will be referred to as Human Relationship Concept.

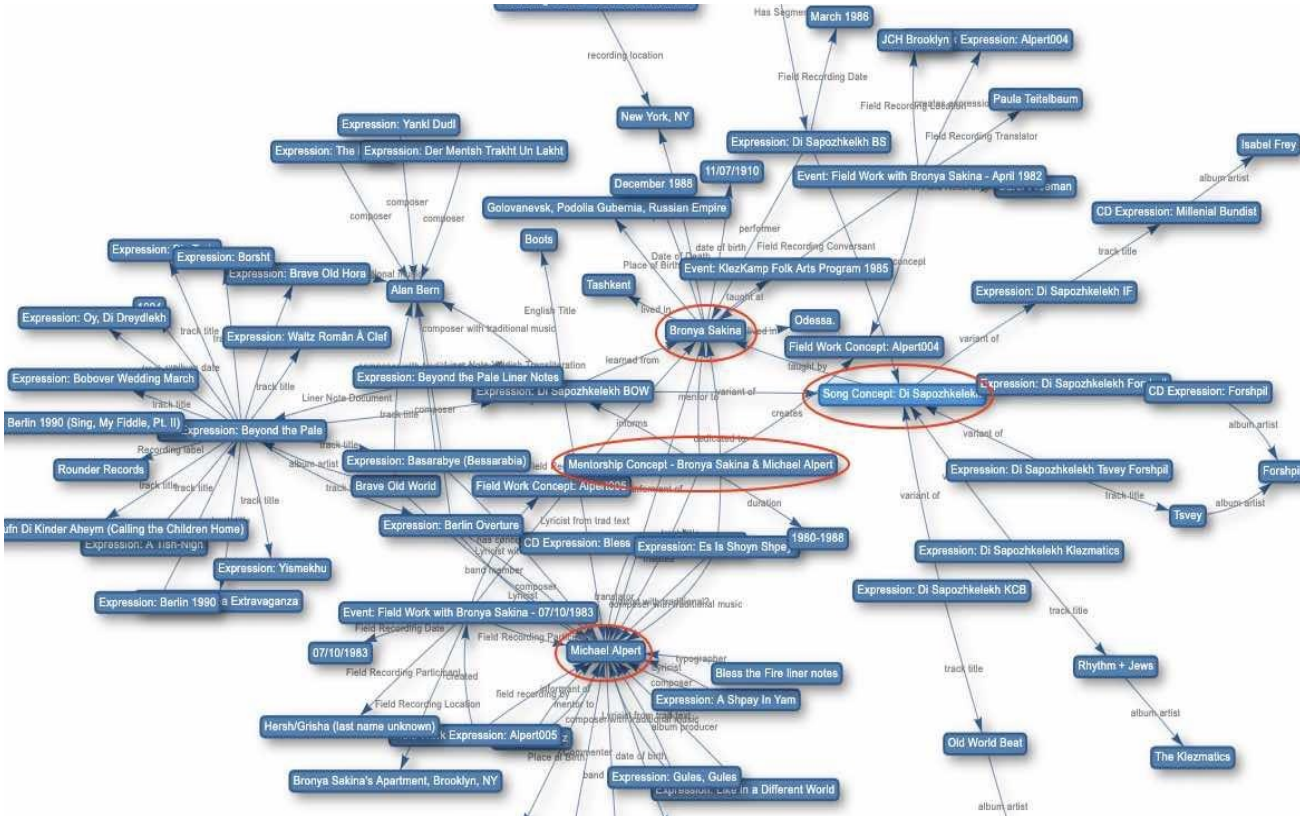

Figure 5. "Documenting Mentorship" prepared for MLA Conference, 2022

and Bronya via a human relationship concept, in this instance a mentorship concept, which provides more context to the relationship. The mentorship concept can be connected to other concepts that include both Bronya and Michael (field work concepts, for example), providing dates, locations, and expressions that tell a more complete story about the people, music, and shared knowledge.

Because the tune and human relationship concepts are an adaptation to the existing DoReMus ontology, the team anticipates the inclusion of these components will not be complex beyond the implementation of the linked open data structure. If successfully paired with powerful search and elegant front-end design inspired by emerging ideas around generous interfaces, these layers of metadata will facilitate discovery in ways never previously possible within the klezmer repertoire.

Appendix I — Knowledge Engines

Knowledge engines are also known as knowledge networks, knowledge graphs, reasoning platforms, reasoning engines, knowledge bases, or first-order logic systems.

Feature	RDF	Property Graph	Knowledge Engine
Represent and query complex/heterogeneous data sets	Yes	Yes	Yes
Add semantics to complex data sets	Yes	Yes	Yes
Perform automated deductive reasoning over large bodies of data / Inference	Yes (with additional tools)	No	Yes
Data Model	Triples	Vertices and edges (usually) without integrity constraints	Assertions, which contain a predicate with any number of arguments, which themselves can be assertions
Supports open world assumption	Nominally yes, in practice no	No	Yes
First-class relations/ Relations between relations (edges of edges)	No (Reification can be used)	No	Yes
Hypergraph (N-ary relations, relationships with multiple arguments)	No	No	Yes
Attribution (edges of edges)	No	No	Yes

Table 1: RDF, property graph and knowledge engine feature comparison.

RDF has limitations that knowledge engines do not, including **attribution**, **validation**, **inference**, and **relations with multiple arguments**. **Attribution** and **validation** refer to information added to relationships (edges) that are not as easily represented as triple stores (e.g. adding a note about who added the data on what date). **Validation** refers to checking or confirming data points, such as confirming the spelling of a particular town name.

The ability to accommodate **relations with multiple arguments** will allow us to store data in a cleaner and more efficient way. Knowledge engines are capable of deriving "facts from facts based on rules," using an approach called **inference**. The ability to enter data about places and their history once and use that data to generate automated tagging for new entries makes for a more complete and less labor-intensive data set.

Examples of how knowledge engine features can be applied to the klezmer corpus:

Because the Klezmer Archive is conceived of as a space with multiple contributors and disparate sources, attribution is of critical importance.

Attribution: Alyson added a connection between Motl Reyder and Dubno, Ukraine on April 27th, 2022. Yiddish expert Sebastian contributed an unusual spelling of a town name.

A data structure that can understand semantic language has the potential to reduce the labor required to input and store information and make that information easier to retrieve in useful and novel ways.

Hypergraph (N-ary relations, relationships with multiple arguments): Triples are fine for modeling simple statements like "Moyshe Beregovsky (subject) was born in (verb) 1874 (object)", but falls short when modeling more complex statements. For example, the sentence "Makenovetsky (argument 1) sent (predicate) his manuscript (argument 2) to Beregovsky (argument 3)" can be represented as a single statement, rather than being forced to turn the sending of the manuscript into a "noun" (i.e. a node) in the data set.

Inference: If the knowledge engine knows that Dubno was a part of the Russian Empire from 1795 to 1918, the Second Polish Republic from 1918 to 1939, and re-occupied by Russia on August 23, 1939, and became part of the Ukrainian SSR in 1944, and part of independent Ukraine in 1991, then if Alyson enters the data point that Freylekhs 1047 was collected in Dubno in 1913, the knowledge engine could tag this entry with both the Russian Empire (time of collection) and Ukraine (present day). When working with transnational musics in eastern Europe, the ability to enter data about places and their history once and use that data to generate automated tagging for new entries makes for a more complete and less labor intensive data set. This is one example of the many possibilities for using knowledge engine reasoning.

Appendix J — Full List of Conference Papers and Public Presentations

- *What's Up With Kiselgof? Introducing KMDMP and the Klezmer Archive* at The Promiscuous World of Jewish Music Series (January 11, 2021)
- *The Crowdsourced Digital Archive* at POLIN Conference: What's New & What's Next, Online (October 5, 2021)
- *What's up with Kiselgof? — KMDMP in Context* at Kahal Braira, Cambridge, MA (November 7, 2021)
- *The Secret Musical Treasures of Ukraine: From Trampled Manuscripts to Digital Collaboration* at Yiddish New York, New York, NY (December 26-29, 2021)
- *The Crowdsourced Digital Archive* at Global Musicology Conference, Online (January 25-27, 2022)
- *Archiving Music Based in Oral Tradition: The Klezmer Archive Project* at Music Library Association's Annual Meeting, Online (March 2-6, 2022)
- *The Kiselgof-Makonovetsky Digital Manuscript Project & the Klezmer Archive: Connecting Cultural Communities to Historical Documents* at Digital Humanities Unbound, Online (May 17-19, 2022)
- *The Crowdsourced Digital Archive* at Society for Ethnomusicology Annual Meeting, New Orleans, Louisiana (November 10, 2022)
- *Klezmer Institute Digital Humanities Projects - KMDMP and KA* for ELTE Folklore Students, Budapest Hungary (November 22, 2022)
- *Archiving Music Based in Oral Tradition: The Klezmer Archive Project* at Digital Humanities Budapest Conference, Budapest, Hungary (November 24, 2022)
- *Using Digital Humanities to Connect the Yiddish Archive to Contemporary Community: The Klezmer Archive Project* at Association for Jewish Studies Conference, Boston, MA (December 18, 2022)
- *Community-Centered Models for Musical Folklore Projects* at SEIF Working Group on Archives Webinar, Online (February 22, 2023)
- *The Klezmer Archive Project: Digital Innovation for Documenting Musics of Oral Tradition* at Répertoire International de Littérature Musicale (RILM) Global Digital Music Studies Conference, Graduate Center at CUNY, New York, New York (April 12-13, 2023)
- *Using Digital Humanities to Connect Archives to Contemporary Community: The Klezmer Archive Project* at Jewish Library Association Conference, Online (June 21, 2023)

Appendix K — Expanded Audience List for KA Tool

This expanded list of potential users of the Klezmer Archive tool was developed during a UX seminar at the first team retreat in October 2021.

Performers: gigging musicians, bandleaders, cantors, dancers, dance Leaders, hobbyist musicians (e.g. Community Bands)

Academics & Researchers: Computational musicologists, Historians, Musicologists/ Ethnomusicologists, Jewish studies, Linguists, Dance scholars, Anthropologists, Non-specialist - world music, geneaologists

Artistic Specialists: composers, choreographers, neo-trad adapters/arrangers, art music professionals, film makers, playwrights, podcast producers, story makers (storytellers, podcasters, theater), GLAM professionals

Teachers & Educators: professors and teachers at all levels, community band leaders, private lesson teachers, workshop leaders, librarians

Jewish Community: synagogue and Jewish community center professionals, Jewish day school teachers and librarians

Students & General Public: music students, college students in Yiddish culture courses, elder hobbyist, non tech-proficient individuals, general interest music listeners