USING WIKIBASE TO LEVERAGE COMMUNITY SOURCED DATA INITIATIVES

Erin Yunes
Virginia Tech

Hollis Wittman
Virginia Tech

James Tuttle
Virginia Tech

Kara Long
Virginia Tech

Abstract

This paper discusses two initiatives of the Data Services unit at the Virginia Tech University Libraries in utilizing Wikibase Cloud to support community-engaged research teams through two distinct projects: The Rematriation Project (REMA) and the Computer Music Preservation Electronic Library (COMPEL). The Rematriation Project focuses on supporting Inuit Arctic communities in digitizing and sharing their cultural, tribal, and scientific knowledge, emphasizing Indigenous Data Sovereignty and community-led decision-making in archival practices. REMA aims to create a culturally reflective digital archive, potentially using Wikibase Cloud, to ensure sustainability and enhance linkages between community generated content. In contrast, the goal of COMPEL is to create a network for exchanging ideas and information among composers, performers, technologists, and scholars of computer music and sound-based digital art. By collecting structured data related to computer music compositions, performances, and technologies, COMPEL seeks to inform preservation strategies and standards specific to this field. Both projects showcase the versatility of Wikibase Cloud in supporting diverse research initiatives and community needs. The work-in-progress reflects on the challenges and surprising synchronicities between projects with highly different data sets.

We will report on the progress and development of two Wikibase projects from two different research teams: The Rematration Project and COMPEL, with support from the Data Services unit in the Virginia Tech University Libraries. The work-in-progress reflects on the challenges and surprising synchronicities between projects with highly different data sets.

Keywords: Wikibase, data modeling, community archives, community data, data governance, computer music, Indigenous data sovereignty, linked data

Introduction

The Data Services unit at the Virginia Tech University Libraries is an interdisciplinary team of librarians and researchers who provide expertise and consulting on designing and operating efficient workflows for data collection, documentation, sharing, and analysis. This team supports faculty, students, and staff in their research efforts, with the goal of sharing and preserving scholarly outputs across all disciplines. As the Data Services team has grown, we have seen an increase in research proposals and projects that require collaborative, community-based and community-facing approaches to data creation, curation, and governance. In 2021, a small team from the Data Services unit began working together to experiment with Wikibase (and, later, Wikibase Cloud) as a tool and platform to support the creation and use of structured data for community-engaged research teams.

In this paper, we compare two long-term projects partnered with the team in Data Services to deploy two separate Wikibase Cloud instances: The Rematriation Project and the Computer Music Preservation Electronic Library (COMPEL). These two projects, as described below, have vastly different audiences, objectives, and goals. However, both teams are closely aligned in the need for digital tools that make it possible for community members to engage with the data, both as co-creators and users.

The Rematriation Project (REMA)

The Rematriation Project (REMA), a collaborative effort between the community organization Aqqaluk Trust of Kotzebue, Alaska, and scholars at Virginia Tech, aims to support Inuit communities in digitizing and making accessible cultural, tribal, and scientific knowledges for their own self-determined needs. At the core of REMA's mission is a commitment to Indigenous Data Sovereignty, ensuring community-first, community-led decision making in the archiving process. This involves a departure from conventional archival methods, focusing instead on collaborative content creation and management practices that honor Iñupiat and Inuit perspectives.

During the first phase of this project, a collection of papers, images, and materials of Caleb Pungowiyi, a noted Siberian Yupik leader and climate activist, was digitized in partnership with the digital imaging lab at Virginia Tech's

University Libraries. This process not only digitally preserved important community documents but also redefined archival practices to prioritize arrangement, cataloging, and metadata creation that reflect Iñupiat and Inuit cultural contexts.

After the pilot collection was digitized, an archival prototype was developed using Obsidian, a note-taking and mind-mapping application. This tool, known for its versatile data handling, was selected for its ability to meet the flexibility and sharing requirements of our team. This prototype allowed for the transformation of Western-centric metadata into terminology and descriptions significant to the community, improving the potential discoverability and interconnection of digitized content for the community. Following the initial data upload, our community partner, Aqqaluk Trust, further refined the data with community-specific terms.

The REMA team is currently developing a Wikibase Cloud instance, based on culturally reflective metadata to create a dynamic, linked digital archive. The new phase of the project focuses on discovering how Wikidata and Wikibase cloud technologies can be used to ensure the archive's sustainability, enhance the linkages in community-generated content, and maintain a knowledge base consistent with Indigenous principles. The Rematriation Project is looking to demonstrate the potential of using freely available softwares and cooperative strategies to leverage Wikibase in supporting community needs as well as serve as a replicable model for similar efforts throughout the Arctic.

The Computer Music Preservation and Electronic Library (COMPEL)

The Computer Music Preservation Electronic Library (COMPEL) seeks to create a network for the exchange of ideas and information among composers, performers, technologists, musicologists, and students of computer music and sound-based digital art. The COMPEL team has three main objectives to support the development of this network: raising awareness and visibility of computer music and sound-based art on the web and in the scholarly community; collecting information related to computer music compositions, performances, hardware and software as structured data that can support complex queries; and, ultimately, to use this structured data to inform and encourage the development of standards and practices for preservation strategies specific to computer music and sound-based digital artworks.

As with all digital media, computer music is a rapidly evolving field and the challenges of preserving both objects and data supersede the pace of development. COMPEL aims to collect data that is currently being lost to create a snapshot of performance and composition culture. The database collects data about individuals, ensembles, individual pieces, and specific performances, utilizing linked data to create a web of connections between them. This provides a centralized, collated location for accessing this data and assists in preserving it for future use. Much of this type of data, particularly performance data, is ephemeral in nature and unlikely to be preserved without intentional effort.

Potential users include (but are not limited to) composers looking for likely performers for their work, performers looking for new pieces to play or composers to work with, and musicologists studying contemporary trends in electronic music. Wikibase Cloud provides a perfect environment in which to explore linked data as a way to explore complex questions about computer music and contemporary performance: a composer could ask if a piece for harp, percussion, and live electronics had ever been written before; a performer could find a piece to play at a concert featuring vibraphone, voice, and triggered fixed media; a researcher could look at how many performances for flute and electronics took place in 2018.

Discussion

REMA and COMPEL initially utilized the same Wikibase instance, with a property that reflected the “VTUL Project” each item or property was affiliated with. These projects were not exclusive to one another; many properties, particularly broader descriptors like “instance of,” applied to both projects. Each project is working on and with its own controlled (more or less) vocabulary, so conflict in word or property use is rare.

Both projects utilize crowdsourced data and user participation to grow and maintain their datasets. For REMA, control and privacy are an integral part of the crowdsourcing process; Wikibase allows these parameters to be put in place. In COMPEL, individual user contributions and edits take priority, as does eventual automation of content addition. Both projects address gaps in the traditionally accepted Wikidata content; the REMA project in its granularity and specificity of archival collections, and COMPEL in its functionally nonexistent notability criteria and low threshold for necessary data to participate. The availability of an open-source linked data platform in which to explore the content of each project has been vital to its success, and we look forward to continuing our work with Wikibase.