Metadata Infrastructure for Sound Recordings

Catherine Lai¹, Ichiro Fujinaga¹, David Descheneau², Michael Frishkopf², Jenn Riley³, Joseph Hafner⁴, and Brian McMillan⁴
Music Technology, Schulich School of Music, McGill University¹; Department of Music, University of Alberta²; Digital Library Program, Indiana University³; McGill University Libraries, McGill University⁴

Introduction
This poster describes the first iteration of a working model for searching heterogeneous distributed metadata repositories for sound recording collections. As the number of digital sound recording collections increase, the task of searching at individual repository sites becomes impractical for end users. For example, users need to:

- Learn multiple interfaces located at different sites
- Manually combine search results
- Meet repeated authentication requirement

The benefits of a distributed search have increasingly become apparent for users because it:

- Searches across multiple repositories
- hides the difference in the mechanics of interactions
- Facilitates the flow of data between applications

System Function
The core function of the metadata infrastructure for sound recordings supports the following processes:

- Search: The Distributed Metadata Service (DMS) supports the searching of metadata repository services, in the context of the federation.
  - All requests intercepted and passed to the Request Router
  - Query requests forwarded to the Request Aggregator
  - Query string synchronously passed to all partner repository interfaces, which are specified in a crosswalk configuration file

Distributed Metadata Services
The Distributed Metadata Services (DMS) perform real-time querying of existing metadata repository services, in the context of the federation:

1. All requests intercepted and passed to the Request Router
2. Query requests forwarded to the Request Aggregator
3. Query string synchronously passed to all partner repository interfaces, which are specified in a crosswalk configuration file

Implementation
Three universities hosting heterogeneous digital repositories of sound recordings participate in the prototype of cross-repository interoperability, each implementing a variety of connection methods and search APIs:

1. Followedrive¹ at University of Alberta, with connection to an RDF store
2. Variations² at Indiana University, with connection to a MySQL database
3. Digital Archive of Handel LP at McGill University, with HTTP connection with custom XML query API

Metadata Integration and Interoperability Issues
Institutions have adopted different metadata models for sound recordings in digital libraries to enhance their music systems. These differences lead to:

- Multiple protocols
- Multiple metadata schemes
- Multiple formats of data

Challenges & Future Work
The absence of uniformity in the repository API and metadata implementation raised issues in areas of search and retrieval, system performance, and presentation. For example:

- Different treatments of Boolean operators, wildcards, operators, and query matching algorithms between the APIs affect the results of federated search operations
- Custom configuration of performance metrics at the individual repositories, which is often implemented and optimized for local use, influences the overall performance
- Incorporate parallel asynchronous access patterns to multiple metadata repositories
- Use a central cache to cache queries and results metadata in an RDF store
- Mapping of custom metadata to different common elements is possible, thus improves the quality of presentation
- Provides visible trace of the query
- Enforce consistent and precise mappings from local to common schemes
- Implement authority control

Acknowledgment
This research is in part funded by the TIST Research Grant of SSHRC, Canada.