

Enriching Metadata for Irish Traditional Music at the American Folklife Center

Introduction

In the past ten years, a number of digital projects worldwide have demonstrated the possibilities afforded by developing semantic ontologies, and some of these frameworks continue to be developed. Linked Jazz, for example, continues to demonstrate how technological discoveries can aid libraries and archives in re-imagining the representation of their holdings that contain music.¹ The recently published Linked Irish Traditional Music project (or LITMUS), a European Horizon 2020 funded project, is further testimony to the support and value for Linked Data in building specific ontologies for orally transmitted music in the archive.² However, experimentation with these resources can reveal both possibilities and limitations for Linked Data when engaged in practice.

Materials and Objectives

Connections in Sound (CIS) was a recent Linked Data pilot study, focused on exploring Irish traditional music in audio collections from the American Folklife Center or (AFC) at the Library of Congress. The project was developed in collaboration with Library of Congress Labs as part of a Kluge Fellowship in Digital Studies and Fulbright Tech Impact award. It sought to develop experimental insight into nascent technologies for this genre. The pilot seeks to contribute to discussions on Linked Data for Irish traditional music, for best practices, methods for analyzing, gathering and representing this particular data. It utilised an ontology developed in the LITMUS project. This ontology is an extension of FRBRoo (Functional Requirements for Bibliographic Records Object Oriented) initiative using SKOS (Simple Knowledge Organization System) in June 2019.³

The CIS team worked closely with reference librarians to identify tunes and songs from fieldworker and recordist documentation in archival collections, transcribing metadata from a range of performance contexts. The process allowed us to test assumptions when connecting common identifiers between descriptions of audio material that had previously existed within separate sections of the archive and the particularities of a range of formats and fieldwork methods from material recorded throughout the twentieth century.

The ontology also facilitated perspectives on the way that tradition is communicated, bringing potential for abundant research in this area and offers insight into advancements for Linked Data with traditional music in general. LITMUS has significantly advanced support for describing this musical genre, including previously unavailable Irish language versions of instruments and tune types.

The development of CIS involved this performer-centred scenario: hundreds of musical pieces from the twentieth century were identified in a variety of formats and regions throughout North America across thirty seven collections with a reference librarian. The performer's need would involve discovering multiple versions of a tune or song. Taking the performer scenario into account, numerous performances of songs and tunes were identified across various collections. Multiple settings of the tune and songs were discovered, with various combinations of performers. These pieces were also found in multiple resources across the World Wide Web. A search was conducted on Library of Congress subject headings and keyword searches of the Library's website. In a sample out of thirty seven identified collections with reference librarians, seventeen collections from the AFC were used. A dataset of over 2500 pieces was gathered from recordists, collector notes, web resources and performances at concerts. It was sorted based on the potential for data identified to be used as Linked Data. Proof of concept Uniform Resource Identifiers (URIs) were developed for each record. RDF Triples were then developed with technologists at the Library.

Working with modern field recordings afforded a number of possibilities for music identification. As an example of possibilities that were encountered, a number of tunes and songs were identified as written by particular composers. They were announced during concert recordings, for example announcements that they would play music from the composer Ed Reavy (Figure 1). These references, even though names were not introduced, could later be identified by researchers and performers interested in finding music by this composer in the collections. This then demonstrates new ways for users of Linked Data to access music that was composed by Ed Reavy, aiding discovery. This was part of a number of possibilities that emerged by interacting in this way with these collections.

LoC URI
<http://id.loc.gov/authorities/names/nr88008912>

WikiData Q number
<https://www.wikidata.org/wiki/Q5335312>

Musicbrainz URI
<https://musicbrainz.org/artist/75490237-7564-4dba-b762-f9d83bcc5630>

Concert Announcement

"I'd like to play a couple of Eddy's [Reavy] tunes..."

Figure 1. Linking concert announcements to Linked Data resources

An important limitation emerged with authority information on the Library website. A lack of comprehensive data contributed to excluding certain performers from possibilities for Linked Data.⁴ In numerous cases it was found that only artists who had produced commercial works were recorded with authority information on the Library's Linked Data resource. This issue affected a significant proportion of records, most obviously fieldwork recordings, and should be considered more fully.

There is a vibrant community in Irish traditional music across Ireland and the United States. Resources have been built by those who have meticulously documented the music, songs and dances, such as IrishTune.info.⁵ In order to reconcile identifiers for musical pieces, this website resource was used.

The screenshot shows the website www.irishtune.info with a navigation bar at the top. The main content area is titled "Irish Traditional Music Tune Index" and features "Tune ID#1712 (P.J. Moloney's)". Below the title is a musical score for a Hornpipe in G Major, 32 bars, with an AABB structure. A table below the score lists recording details:

Rhythm	Bars	8-bar phrase structure	Mode
Hornpipe	32	AABB	G Major

Below the table, there is a section for "Titles" and a "Discography" section. The discography section contains a table of recordings:

Year Recorded	Track #	[Album code] Artist. Title (Link to Album Info page). Performers (instruments).
~1958	37#2	[TDMI] Traditional Dance Music of Ireland. The Jimmy Hogan Trio.
~1959	3#3	[SRPJM] Seán Ryan, P.J. Moloney. Traditional Music of Ireland. Volume One. Digitally Remastered. Seán Ryan (fiddle), P.J. Moloney (flute).
~1974	A.6#1	[McG/Cool] McGreevy and Cooley. McGreevy and Cooley. John McGreevy (fiddle), Seamus Cooley (flute). For this tune: Seamus (flute).
~1996	8#2	[ST ES] Seamus Tansley. Easter Snow. Seamus Tansley (flute).
~1996	12#1	[JB 4] Joe Burke, Charlie Lennon. The Bucks of Oranmore. Joe Burke (accordion), Charlie Lennon (piano).
~2002	3#1	[BRny 2] Brian Rooney. Leitrim to London. Brian Rooney (fiddle, accordion).

Figure 2. The website www.irishtune.info which records URL IDs for music, but not URIs

There are also multiple other URI resources on the Web that could link to the items within the AFC, such as Musicbrainz.org. However, there is a distinct lack of WikiData.org entries for cross-referencing songs and tunes and no central resource, even though many resources have been created for learners. In addition to this, most resources such as thesession.org and irishtune.info have been created with the needs of performers in focus. As a result, URIs do not exist on these websites.

The CIS project was also extended to include transcribed story metadata. In this musical genre, stories from performers add a vital understanding about the context surrounding a tune or song. A Linked Data ontology associated with the American Folklore Society Ethnographic Thesaurus was used to build upon these possibilities for describing stories. As a result of this effort, multiple types of connections were made between these tunes and songs. These triples were tested using SPARQL-visualizer, developed by Mads Holten.⁶

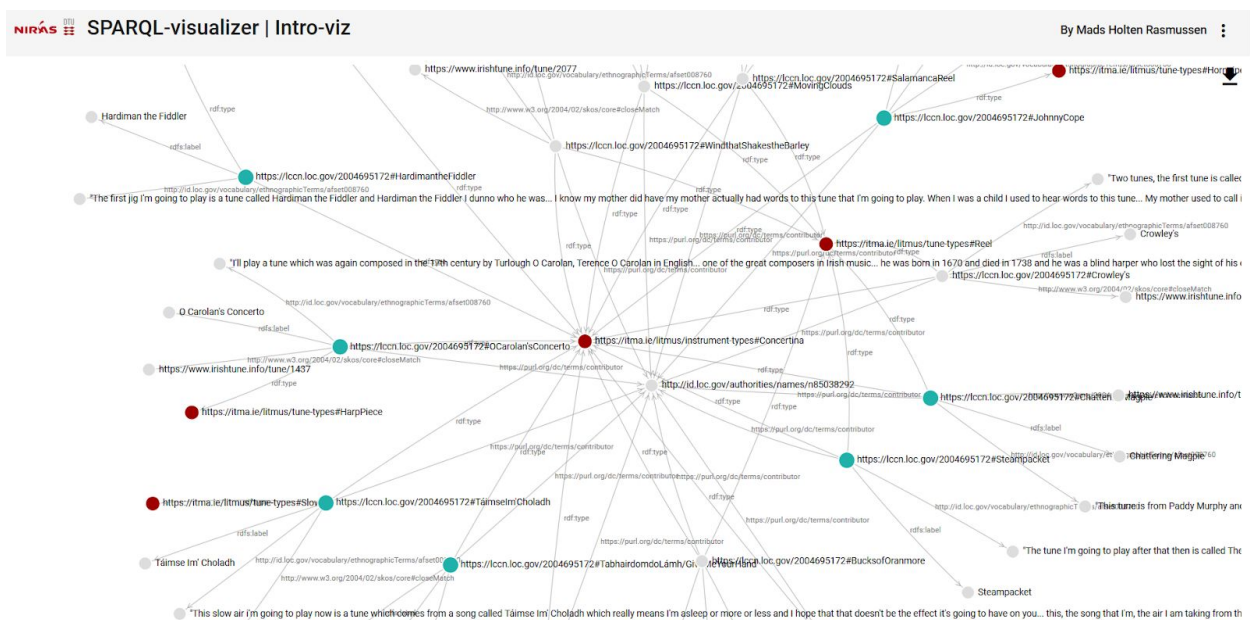


Figure 3. SPARQL-visualizer by Mads Holten

These resources were then linked to resources on the World Wide Web such as the Séamus Connolly Collection located in archives at Boston College. This collection has a number of datasets that were usable, so that SPARQL queries could be set up. Queries were constructed within the website Data.world (Figure 4).

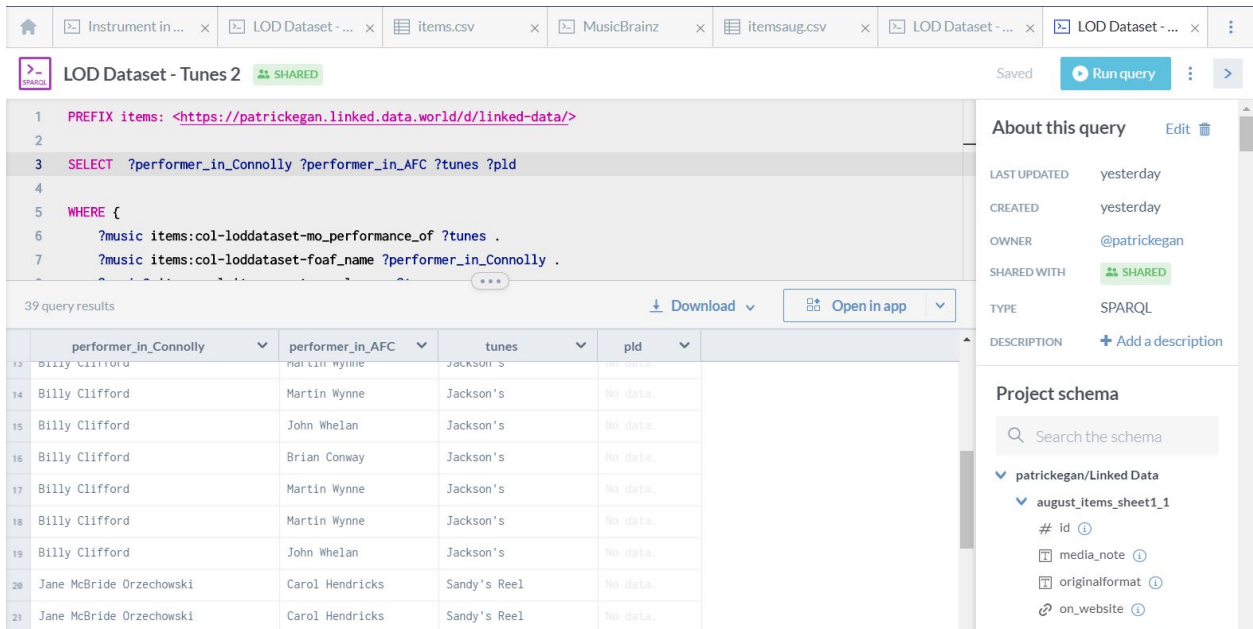


Figure 4. Data.world representation of connecting triples to Séamus Connolly Collection

In order to demonstrate connections between audio collections at the Library of Congress an application was developed in Glitch (Figure 5). The application reads the most popular tunes from the dataset and displays the location where they were recorded during the twentieth century. Visitors can experience the same tune across different collections from the AFC from a number of eras and formats.

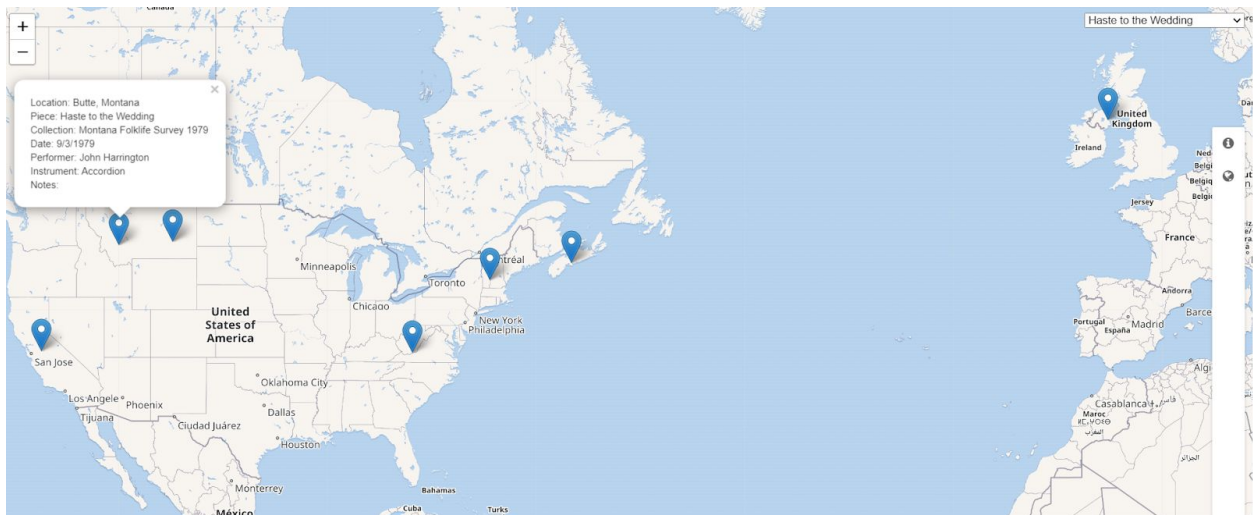


Figure 5. Glitch application representing the most identified tunes within AFC audio collections <http://rural-satellite.glitch.me>

Construction of the CIS dataset followed Force11 FAIR Guiding principles, documenting all interactions during the building of datasets and code.⁷ Archival material has been

made findable, accessible, interoperable and replicable on the web, and data can be analysed.⁸ The resultant dataset can be used by others, for example through future development.

Conclusion

The collections of audio pertaining to Irish traditional music at the American Folklife Center demonstrate a wide diversity of audio formats, performance contexts and metadata. In attempting to utilise Linked Data to connect material, affordances were discovered that had not been possible in previous research. There are also a number of limitations with regard to the current infrastructure for Irish traditional music on the World Wide Web, which highlight the focus of authority information is generally attributed to more well known artists in standard setting organisations such as the Library of Congress. As a result, the development of Linked Data resources poses many questions not only for technologists and archivists, but prompt further consideration by scholars for the range of users who seek to engage with archival material. Researchers can play a key role in the development of a more balanced representation of Irish traditional music for Linked Data.

There are multiple possibilities for future projects, and the dataset and tools discussed here continue to be developed. Future applications will establish connections with repositories around the world. This project would also benefit greatly from the integration with WikiData.org. Doing so would enable the use of tunes and songs as identifiers, establishing a central authoritative resource and expanding beyond current infrastructure.

Notes

1. See: Pattuelli, M. Cristina, Karen Hwang, and Matthew Miller. "Accidental discovery, intentional inquiry: Leveraging linked data to uncover the women of jazz." *Digital scholarship in the humanities* 32.4 (2017): 918-924.
2. For more on LITMUS, see: Weissenberger, Lynnsey K. "Stories, songs, steps, and tunes: A linked data ontology for Irish traditional music and dance." *International Society for Knowledge Organization, UK/Ireland Chapter, 2017 Annual Conference, Knowledge Organisation: What's the Story? 11-12 September, 2017, London, United Kingdom*. 2017.
3. For more about FRBRoo and SKOS implementation in LITMUS, see: "LITMUS Project Information." *ITMA*, 4 Apr. 2020, www.itma.ie/litmus/info.
4. Authority information was accessed through python scripts but also manually. Library of congress Linked Data service available here: "Home - LC Linked Data Service: Authorities and Vocabularies | Library of Congress." *Home - LC Linked Data Service: Authorities and Vocabularies | Library of Congress*, Library of Congress, id.loc.gov/. Accessed 19 July 2020.
5. See: Ng, Alan. "Irishtune.info Irish Traditional Music Tune Index." *Irishtune.info Irish Traditional Music Tune Index*, Alan Ng, www.irishtune.info. Accessed 19 July 2020.
6. For more about SPARQL-visualizer, see "MadsHolten/sparql-visualizer." *GitHub*, 26 Mar. 2020, github.com/MadsHolten/sparql-visualizer. Accessed 19 July 2020.
7. For more about FAIR Guiding Principles, see: Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1 (2016): 1-9.
8. Resultant datasets and code will be available on both GitHub and Glitch: For more about the data and code used in this project, see: Egan, P. "Rootseire - Overview." *GitHub*, 16 May 2019, github.com/rootseire. Accessed 21 June 2020.