2025 © Caleb Triscari, *Context* 50 (2024): 109–22.

Research Report

Considerations for Future-Proofing Digital First Nations Music Collections

Caleb Triscari

Researchers, when searching for musical material through an online catalogue, may scarcely consider how the resources included in the search results came to be suggested by the catalogue system. A catalogue's search results are dependent on the metadata that describes the resources held within that collection,[1] which can be stored in a variety of formats, standardised or unstandardised, each with their own benefits and limitations with respect to discoverability and interoperability with other systems. Music originating from First Nations communities around Australia often requires unique contextual knowledge in order to be discovered easily in a collection. Any First Nations music recordings that are held in a digital archive would ideally require that contextual knowledge to be stored as metadata, alongside the recording: this would ensure access conditions are respected (where appropriate), and information on which First Nations communities are involved is readily accessible.

Several systems seek to ensure this accessibility in a culturally safe way, one being the Connecting Collections and Communities Tool (hereafter CCCT), developed by Treloyn et al. in

[1] Metadata is defined as 'structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource'. See Jenn Riley, *Understanding Metadata* (Bethesda, MD: NISO Press, 2004), 1.

collaboration with First Nations users and organisations.[2] A relational database template built with proprietary FileMaker Pro software,[3] the CCCT seeks to act as a resource-rich database schema for First Nations song material, which can be used to ensure specific metadata is stored for accessibility and long-term preservation purposes. The database pursues the goals of Indigenous data sovereignty by providing First Nations communities with a tool to manage records relating to their cultural heritage,[4] and has been deployed locally and internationally to date, including in the Kimberley region and Uganda, with promising results.[5]

The CCCT's database schema—that is, the fields of each relation (or table) in the database and any related constraints—is bespoke, to account for the nature of First Nations musical systems, and does not adhere to any recognised standard due to the complex metadata it is designed to handle.[6] This will raise future issues when making the CCCT interoperable with other systems: an unstandardised metadata schema will conflict with other systems and make it difficult to collaborate with other collections.

These challenges were the focus of my recent research project, which proposed a collection of standards, ontologies, and thesauri to which the CCCT metadata schema can be mapped (or 'crosswalked'), in order to improve its accessibility and interoperability.[7] In proposing these standards, my project closely examined the needs of the database, technical documentation on existing standards, and relevant academic literature. It was carried out with the intent of ensuring the database and the metadata it housed were as 'future-proof' as possible. This meant looking toward emerging data structures with promising benefits, such as linked data (explained later in this article). However, it became clear through this project that the metadata standards used presently by collecting institutions are still insufficient in accommodating the complexities of metadata intrinsic to First Nations music material: more research will be required if the accurate description and accessibility of First Nations music material is to be improved.

From a global historical perspective, issues of accessibility in sound recording collections have existed ever since the format rose to prominence in the nineteenth century.[8] Large

[2] Sally Treloyn et al., 'Connecting Communities and Collections Tool (CCCT),' figshare, 2024, https://doi.org/10.6084/M9.FIGSHARE.13396310.V4.

[3] 'Start,' Connecting Collections and Communities Tool, accessed 9 June 2024, https://wiki.ruiac.org/doku.php?id=start. Developers of the CCCT have expressed their eventual intent to migrate the schema to an open-source equivalent.

[4] Indigenous data sovereignty refers to the right of Indigenous peoples to exercise ownership, collection and application over information or knowledge that concerns them, their culture or land, in any format or medium. See Tahu Kukutai and John Taylor, 'Data Sovereignty for Indigenous Peoples: Current Practice and Future Needs,' in *Indigenous Data Sovereignty: Toward An Agenda*, ed. Tahu Kukutai and John Taylor (Canberra: ANU Press, 2016), https://doi.org/10.22459/CAEPR38.11.2016.01.

[5] Erin Parke, 'Rare Captain Cook Dance among Hundreds Revived in the Kimberley,' *ABC News*, 12 July 2021, https://www.abc.net.au/news/2021-07-12/captain-cook-yagan-dance/100284908.

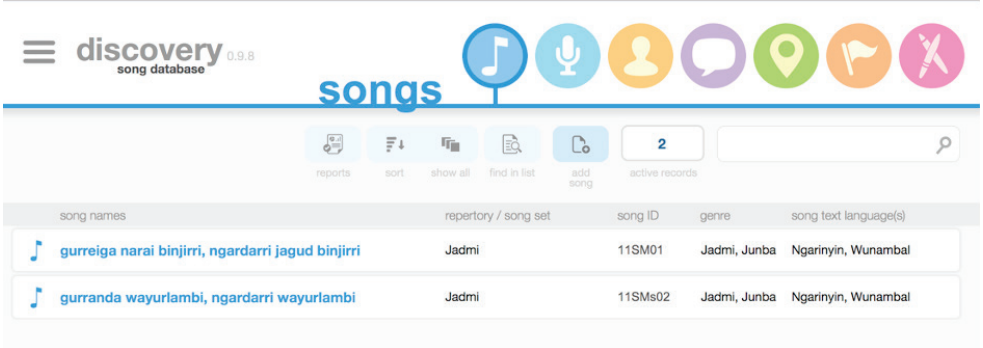
[6] Sally Treloyn, Rona Goonginda Charles, and Pete Myadooma O'Connor, 'Dancing with the Devil (Spirit): How Audiovisual Collections Reveal and Enact Social and Political Agency in Dance and Song (A Case from the Kimberley),' *Preservation, Digital Technology & Culture* 50.3–4 (2021): 117–29, https://doi.org/10.1515/pdtc-2021-0027.

[7] A thesaurus is 'a controlled vocabulary arranged in a specific order and characterised by three relationships: equivalence, hierarchical and associative,' while an ontology is a 'formal, machine-readable specification … in which concepts, properties, relationships, constraints and axioms are all explicitly defined.' See Patricia Harpring, *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*, 1st ed. (Los Angeles: Getty Research Institute, 2010), 228, 237.

[8] C. Rockelle Strader,'Cataloging Music Sound Recordings in the United States: An Evolution of Practice and Standards,' *Notes* 72.2 (2015): 276–327, https://doi.org/10.1353/not.2015.0150.

amounts of musical material are still inaccessible to their communities of origin, for several reasons including outdated audio formats, insufficient descriptive metadata, and ineffective software. To address some of these issues, my research assists in facilitating broader access, within communities of origin, to First Nations musical recordings by proposing a standardised schema to which the current metadata schema can be crosswalked. Further, this study will help improve the database's interoperability, allowing the metadata to be understood by other systems using the same metadata standard. Through this proposed crosswalk, metadata recorded in the CCCT database will theoretically have a broader reach and improve accessibility. This study's findings and outputs contribute to the wider interdisciplinary academic fields of applied ethnomusicology, archival science, and information science—which share a focus on the sustainable repatriation and improved accessibility of sound recordings.

Figure 1. Screenshot of the CCCT database displaying entries in the Songs relation.



Conventions

To assist in differentiating between technical and non-technical terminology, a number of typographical conventions have been used in this report:

- The elements included within an ontology are divided into two categories: classes and properties. Classes are used to assign categories or types to a resource (for example, `Person`, `Place`, `Event`, `MusicComposition`), whereas properties are used to assign additional information to the resource (for example, `name`, `alternateName`, `description`, `subjectOf`, `sameAs`).
- The font `Courier New` is used to distinguish data and code snippets, and RDF expressions from the rest of the report text.
- Classes and properties belonging to a schema or ontology will frequently be expressed in a shorthand form to appear cleaner throughout this report. For example, the unique identifier https://schema.org/MusicComposition will be shortened to schema:MusicComposition.
- Classes are capitalised. For example, `MusicComposition`, `AudioObject` or `Comment`.
- Properties are written in lower camel case. For example, `rightsHolder`, `encodingFormat` or `identifier`.
- Names of relations (or tables) within the CCCT database are capitalised, while the fields within relations are wrapped in quotation marks. For example, the Songs relation has 'name', 'genre', 'composer' and 'description' fields.

Literature Review

Before devising the crosswalk, a literature review was conducted to assess recent research in this multidisciplinary field. In this context, it assessed common standards, ontologies, and thesauri used to record music metadata in collections, in addition to relevant academic literature in musicology, ethnomusicology, information science and archival science with respect to First Nations music collections. The intention was to outline the importance of conducting this research and to explore the benefits and limitations of certain standards that were considered for the project. It became evident that no single standard could comprehensively accommodate the complexities of First Nations music metadata, and the issues of accessibility, interoperability, and future-proofing archives persist.

Accessibility and interoperability within digital collections have been discussed extensively, and many issues remain unresolved. Archives, and the standards they employ, struggle to future-proof their systems, particularly those for items of non-Western origin. In 2004, Barwick described the advancements in digital systems that allowed for the Pacific and Regional Archive of Digital Sources in Endangered Cultures (PARADISEC) to develop as a distributed archive that broke down geographic or technological barriers to the communities that have material in the archive.[9] In the decades since Barwick's article, new digital systems have emerged with the purpose of facilitating research and connecting material with their communities of origin. The Global Jukebox, for example, is an online resource for exploring music from cultures around the world, with references to historical, geographic, linguistic, and ethnographic contexts.[10] In Australia, Treloyn and Emberley detail the process of establishing a local iTunes database in the Kimberley region to repatriate Ngarinyin, Worrorra and Wunambal dance-song recordings, complete with complex metadata including the composer, their clan of origin, living family, and clan members.[11] Elsewhere, Thorpe et al. recount the process of developing Mukurtu, an online content management system that empowers First Nations communities to have greater governance over their digital heritage material, including access protocols specific to First Nations heritage material.[12] What is evident is that each system has unique requirements in order for its material to be effectively described through metadata. Standards handle this complex metadata in unique ways, albeit with varied success. The suitability of some standards to fit within this CCCT migration project are discussed below.

Standards: Towards RO-Crate

As mentioned earlier, the literature review also assessed the benefits and limitations of commonly used metadata standards that are being considered for the purposes of our study. There is a lack of literature on standards, future-proofing, and interoperability with respect to

---

[9] Linda Barwick, 'Turning it all Upside Down ... Imagining a Distributed Digital Audiovisual Archive,' *Literary and Linguistic Computing* 19.3 (2004): 253–63, https://doi.org/10.1093/llc/19.3.253.

[10] Anna L. C. Wood et al., 'The Global Jukebox: A Public Database of Performing Arts and Culture,' ed. Steven R. Livingstone, *PLOS ONE* 17.11 (2022): e0275469, https://doi.org/10.1371/journal.pone.0275469.

[11] Sally Treloyn and Andrea Emberly, 'Sustaining Traditions: Ethnomusicological Collections, Access and Sustainability in Australia,' *Musicology Australia* 35.2 (2013): 159–77, https://doi.org/10.1080/08145857.2013.844473.

[12] Kirsten Thorpe et al., 'Designing Archival Information Systems through Partnerships with Indigenous Communities: Developing the Mukurtu Hubs and Spokes Model in Australia,' *Australasian Journal of Information Systems* 25 (2021), https://doi.org/10.3127/ajis.v25i0.2917.

ethnomusicological archives. However, this scope could be widened to incorporate literature that discusses both linguistics-oriented archives holding ethnomusicological material and the standards they employ, such as Dublin Core, OLAC, and RO-Crate.

Dublin Core and OLAC

Several archives around the world focusing on the linguistic and musical aspects of their material use the Open Language Archives Community (OLAC) standard for their metadata.[13] The OLAC standard is built on the Dublin Core, a highly popular set of fifteen elements used in a variety of archival contexts.[14] Dublin Core elements include `title`, `author`, `creator`, `description` and `language`. A supplementary set of elements, known as the DCMI Terms, were subsequently introduced and include additional terms such as `accessRights`, `contributor`, `rightsHolder` and `publisher`.[15] The OLAC standard builds on the Dublin Core and DCMI Terms by adding qualifying details about the roles of people involved in creating and preserving the material or the languages featured.[16] For instance, the compiler element from the OLAC Roles vocabulary can extend the broad Dublin Core `contributor` term (see Fig. 2). This added detail is instrumental to providing further context to a recording's origin.

Figure 2. An excerpt of a PARADISEC record using the OLAC standard. It assigns linguist Gavan Breen the role of compiler; 'Metadata Record for "Notes on Arrernte Vowels,"' PARADISEC, accessed June 9 2024, https://catalog.paradisec.org.au/oai/item?verb=GetRecord&identifier=oai:paradisec.org.au:GB38-001&metadataPrefix=olac.

```
<dc:contributor xsi:type="olac:role" olac:code="compiler">Gavan Breen</dc:contributor>
```

OLAC's specificity, combined with Dublin Core's popularity, lends itself to being a prime contender for inclusion in the CCCT crosswalk. Dublin Core is commended for its capacity to flexibly describe a wide range of resources with such a simple schema.[17] However, it is not without limitations. Even with added descriptive quality from OLAC, the Dublin Core's primary criticism is that its broad set of elements is far too simple to accurately describe resources in a collection.[18] The First Nations recordings stored in the CCCT database need rich, descriptive metadata in order to make them accessible to their communities of origin, such as information on the traditional owners of a certain geographic region. Dublin Core's set of metadata elements will not be able to effectively describe this. Toner has previously criticised the effectiveness of Dublin Core for storing metadata of Indigenous material, arguing that it is 'based on Western systems of knowledge management' and more must be done to accommodate

[13] Steven Bird and Gary Simons, 'White Paper on Establishing an Infrastructure for Open Language Archiving,' Open Language Archives Community, 7 December 2000, http://www.language-archives.org/documents/white-paper.html.

[14] DCMI Usage Board, 'Dublin Core™ Metadata Element Set, Version 1.1: Reference Description,' DCMI, 14 June 2012, https://www.dublincore.org/specifications/dublin-core/dces.

[15] DCMI Usage Board, 'DCMI Metadata Terms,' DCMI, 20 January 2020, https://www.dublincore.org/specifications/dublin-core/dcmi-terms.

[16] 'OLAC Role Vocabulary,' Open Language Archives Community, 6 April 2006, http://www.language-archives.org/REC/role.html.

[17] Carolyn Guinchard, 'Dublin Core Use in Libraries: A Survey,' *OCLC Systems & Services: International Digital Library Perspectives* 18.1 (2002): 40–50, https://doi.org/10.1108/10650750210418190.

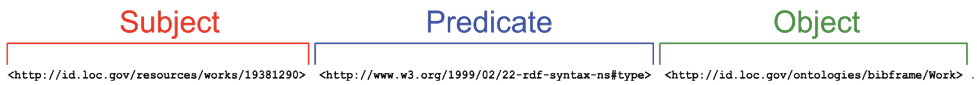
[18] Corey Harper, 'Dublin Core Metadata Initiative: Beyond the Element Set,' *Information Standards Quarterly* 22.1 (2010): 22, https://doi.org/10.3789/isqv22n1.201004.

other types of knowledge.[19] Further, OLAC's controlled vocabulary of roles does not include terms featured in the CCCT database, such as sound engineer or dancer (discussed later in this article). OLAC and Dublin Core constrain the metadata's specificity, but alternatives are available, as explained below.

Linked Data

Linked data is a relatively new technology that was designed to make the vast amount of information on the web machine-readable and traversable. Proposed by Berners-Lee et al., linked data is a form of structured data capable of linking multiple independent datasets to establish an enriched graph of information.[20] This is done through a series of shared ontologies and unique identifiers, often taking the form of a URL. The most common standard used in linked data is the Resource Description Framework (RDF). Data expressed through the RDF standard can exist in many formats, but it ultimately comprises three components: a subject, predicate, and object. This 'triple' is the foundation of every connection in a linked dataset (see Fig. 3).

Figure 3. An RDF triple from a Library of Congress metadata record for a Passamaquoddy song. It declares that the resource (subject) has been assigned the type (predicate) work (object); 'Metadata Record for "Passamaquoddy Song of the Snake Dance" [N-Triples],' Library of Congress, accessed June 9 2024, https://id.loc.gov/resources/works/19381290.nt.



Linked data has been described as 'schema neutral' because it can flexibly establish relationships between any two resources without overarching constraints.[21] This quality was beneficial to Corn and Patrick's research that sought to map Yolŋu and Warlpiri kinship protocols to machine-readable ontologies.[22] It has also been useful in library settings and adopted by the Library of Congress, Bibliothèque nationale de France, Biblioteca Nacional de España, and British National Bibliography.

Despite many of the issues linked data could potentially resolve, its uptake in libraries and archives has been slow outside of major institutions with sufficient resourcing.[23] Linked data also represents a radical shift from the major metadata standards that have been used in collections for decades, such as Machine Readable Cataloguing (MARC) in libraries, which may explain some of the hesitation.

---

[19] Peter Toner, 'History, Memory and Music: The Repatriation of Digital Audio to Yolngu Communities, or, Memory as Metadata,' in *Researchers, Communities, Institutions, Sound Recordings*, ed. Linda Barwick et al. (Sydney: University of Sydney, 2003), 14, http://hdl.handle.net/2123/1518.

[20] Tim Berners-Lee, James Hendler, and Ora Lassila, 'The Semantic Web,' *Scientific American* 284.5 (2001): 34–43; Tim Berners-Lee, 'Linked Data - Design Issues,' World Wide Web Consortium (W3C), 27 July 2006, https://www.w3.org/DesignIssues/LinkedData.html.

[21] Seth van Hooland and Ruben Verborgh, *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata* (London: Facet Publishing, 2014), 44.

[22] Aaron Corn and Steven Wantarri Jampijinpa Patrick, 'Exploring the Applicability of the Semantic Web for Discovering and Navigating Australian Indigenous Knowledge Resources,' *Archives and Manuscripts* 47.1 (2019): 131–52, https://doi.org/10.1080/01576895.2019.1575248.

[23] David Stuart, *Practical Ontologies for Information Professionals* (London: Facet Publishing, 2016), 28.

RO-Crate

Research Object-Crate (RO-Crate) is a new standard that draws on the linked data concept to store the metadata of research data and other collected material.[24] Soiland-Reyes et al. proposed RO-Crate in response to the growing expectation of research data to be made available to other researchers, and to adhere to the Findable, Accessible, Interoperable, and Resuable (FAIR) principles of data management.[25] RO-Crate is stored as JSON-LD, which takes advantage of the popular JavaScript Object Notation (JSON) data format and incorporates RDF elements to give it linked data functionality. Specifically, RO-Crate prioritises use of the Schema.org ontology for its linked data functionality, though other ontologies may also be used.[26] Schema.org is a deeply expansive ontology comprising more than 800 classes and 1,400 properties, used to create structured data on the internet. Schema.org has been used previously to effectively describe music and other cultural heritage metadata, mainly in European collections.[27]

RO-Crate has already been adopted by a small number of research projects in Australia. PARADISEC, for instance, recently converted its catalogue to the RO-Crate format in an attempt to co-locate their metadata on the same disk as the material, which is not always possible when using proprietary databases (see Fig. 4).[28] PARADISEC also uses the RO-Crate standard when repatriating digital audiovisual material to communities. Using hard disks, flash drives or Raspberry Pi devices, the PARADISEC team package together audiovisual material with its accompanying metadata to create a bespoke catalogue of downloaded material from the archive. This is highly beneficial in geographic locations where internet access is scarce.[29]

In the context of the CCCT, migrating metadata out of the proprietary Filemaker Pro software is a primary goal of the CCCT team. Further, being able to repatriate collection items when requested is a longer-term aim, and it would be beneficial to adopt a standard that can provide for this in the future. Both needs appear to be supported by the features provided by RO-Crate. After reviewing the literature, it became clear that RO-Crate is a suitable standard to adopt for the purposes of this project. It fulfils the criteria of the project's aims, which is to find an open-source solution that appropriately describes the complex metadata featured in the collection. The following section outlines how fields from the CCCT relational database schema were crosswalked to linked data ontologies for use with RO-Crate. Developing a crosswalk involved selecting supplementary ontologies and conceptualising how pieces of metadata can connect to each other through the use of blank nodes.

---

[24] Stian Soiland-Reyes et al., 'Packaging Research Artefacts with RO-Crate,' *Data Science* 5.2 (2022): 97–138, https://doi.org/10.3233/DS-210053.

[25] Mark D. Wilkinson et al., 'The FAIR Guiding Principles for Scientific Data Management and Stewardship,' *Scientific Data* 3.1 (2016): 160018, https://doi.org/10.1038/sdata.2016.18.

[26] 'Schema.Org,' Schema.org, accessed 9 June 2024, https://schema.org.

[27] Nuno Freire, Valentine Charles, and Antoine Isaac,'Evaluation of Schema.Org for Aggregation of Cultural Heritage Metadata,' in *The Semantic Web*, ed. Aldo Gangemi et al., vol. 10843, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2018), 225–39, https://doi.org/10.1007/978-3-319-93417-4_15; Anna Neovesky and Frederic Von Vlahovits, 'Interconnecting Music Repositories with Semantic Web Technologies—An RDF- and Schema.Org-Based Approach,' *Digital Scholarship in the Humanities* 36, Supp. 1 (2021): 49–54, https://doi.org/10.1093/llc/fqaa019.

[28] Nick Thieberger, 'Doing It for Ourselves: The New Archive Built by and Responsive to the Researcher,' *Digital Humanities Quarterly* 17.1 (2023): 6.

[29] See Nick Thieberger, 'Running Repairs on a Raspberry Pi in the Field,' PARADISEC, 21 February 2024, https://www.paradisec.org.au/blog/2024/01/running-repairs-on-a-raspberry-pi-in-the-field.

Figure 4. An excerpt of metadata in a PARADISEC record stored in the RO-Crate format; 'Catalogue Record for "Manuel Wayane Reading Words,"' PARADISEC, accessed 9 June 2024, https://mod.paradisec.org.au/view/NT5/200801.


```
"contentLanguages": [
  {
    "@id": "#language_erk",
    "@type": "Language",
    "code": "erk",
    "location": {
      "@id": "#geo_168.159,-17.83,168.594,-17.585"
    },
    "name": "Efate, South"
  }
],
"countries": [
  {
    "@id": "#country_Vanuatu",
    "@type": "Country",
    "code": "VU",
    "name": "Vanuatu"
  }
],
"digitisedOn": "Tue Jul 22 2008 14:00:00 GMT+0000 (Coordinated Universal Time)",
"external": 0,
"languageAsGiven": "Nafsan",
"metadataExportable": 1,
"originalMedia": "Zoom H2 recorder, internal mic",
"originatedOn": "2008-07-14",
```


Crosswalking Process

It is clear from the CCCT's unique and unstandardised schema that the Schema.org ontology, preferred by RO-Crate, will not completely satisfy the requirements of crosswalking the metadata. Therefore, additional ontologies alongside Schema.org will need to be used. The intention of this research project was to adopt pre-existing reputable ontologies, where possible, to prioritise interoperability with other systems,[30] however, as previously discussed, there were occasions where this was not possible. In addition to Schema.org, the other ontologies included in this project are outlined in Table 1.

After selecting the ontologies, the crosswalking process involved working through each field in a relation within the CCCT, and identifying the most appropriate equivalent property among the selected ontologies. Due to the limited timeframe to complete this project, only the most important ten relations were crosswalked: Songs, Recordings, Collections, Events, People, Segments, Instruments, Languages, Places, and Notation. This process had varying degrees of success. To assist with the crosswalking process, a sample CCCT metadata record was provided by Treloyn. The record focused on the musical work of Scotty 'Nyalgodi' Martin, a Ngarinyin elder and composer and performer of the dance-song genre Junba.[31]

A large portion of the metadata stored in the CCCT can also be found in conventional archive systems, including the work's title, its creator(s), date of creation, and format. Schema.org

[30] Stuart, *Practical Ontologies*, 79.

[31] For more information on Scotty Martin, see Sally Treloyn, 'Scotty Martin's Jadmi Junba: A Song Series from the Kimberley Region of Northwest Australia,' *Oceania* 73.3 (2003): 208–20.

Table 1. A list of ontologies included in the project's proposed crosswalk.

Ontology	Description
Resource Description Framework Concepts Vocabulary	A fundamental ontology used to define basic concepts in linked data and resource description
Simple Knowledge Organisation System (SKOS)	A fundamental ontology used to establish alternate and preferred names for resources, and link broader and narrower concepts (e.g. percussion is broader than clapsticks)
Music Ontology	Used mainly when discussing instruments, as Schema.org does not have a musical instruments class or related properties
Dublin Core Metadata Initiative (DCMI) Terms	A popular ontology, used in this instance when no other suitable alternative was available
Bibframe	Developed by the Library of Congress. Primarily used in this project to attribute an individual's contribution to a song, recording or collection.

Compiled from: DCMI Usage Board, 'DCMI Metadata Terms,' DCMI, January 20 2020, https://www.dublincore.org/specifications/dublin-core/dcmi-terms; Library of Congress, "BIBFRAME 2.0," Library of Congress, accessed November 17 2024, https://id.loc.gov/ontologies/bibframe.html; Yves Raimond et al., 'Music Ontology,' Music Ontology, July 22 2013, http://purl.org/ontology/mo; World Wide Web Consortium, 'SKOS Simple Knowledge Organization System Reference,' World Wide Web Consortium (W3C), August 18 2009, https://www.w3.org/TR/skos-reference; World Wide Web Consortium, 'RDF Schema 1.1,' World Wide Web Consortium (W3C), February 25 2014, https://www.w3.org/TR/rdf-schema.

has addressed these basic metadata requirements by creating properties that can be used for those purposes. For example, the 'name' field found in several relations within the CCCT can simply be mapped to the URL http://schema.org/name, which represents the `name` property from the Schema.org ontology.[32]

Table 2. A partial crosswalk demonstrating the terms used in the CCCT and a proposed equivalent.

Term used in CCCT schema	Proposed linked data equivalent
'name'	`schema:name`
'type'	`rdf:type` or `@type`
'composer'	`schema:composer`
'duration'	`schema:duration`
'description'	`schema:description`
'region'	`schema:containedInPlace`
'genre'	`schema:genre`
'nameAKA'	`schema:alternateName`

Blank Nodes

Whenever a link is formed between two items in different relations in the CCCT, a unique identifier is given to that link. For example, an identifier would be created for the link made between Person A and Recording A, in which Person A plays the clapsticks in Recording A. These multifaceted connections are difficult to express in just one subject-predicate-object

[32] 'Name - Schema.Org Property,' Schema.org, accessed 9 June 2024, https://schema.org/name.

triple, as they include more than three components (that is, a person, their contribution to the recording, the role they play in the contribution, and the recording itself).

To resolve this, the Resource Description Framework (RDF) standard for linked data permits 'blank nodes', which are local, non-persistent and non-unique identifiers used to connect atomic components of a dataset together.[33] Figure 5 is a hypothetical excerpt of metadata from the CCCT record provided for the purposes of this project, stored in JSON-LD, the same format as the proposed RO-Crate standard. The recording lists one contribution, which is assigned a blank node identifier '_:N602af0ef67c5442b8c23e72451c09739'. The blank node then links to an example identifier given to Scotty Martin, as well as identifiers for the roles he played in the recording: singer, instrumentalist, and copyright holder. It also lists 'clapstick player' as a role, however an identifier has not been assigned for this role, as will be explained below. This hypothetical example shows how blank nodes can be applied when crosswalking the various contributions made to songs, recordings, or events stored in the CCCT database.

Figure 5. A crosswalked CCCT record demonstrating the roles assigned to Scotty Martin's contribution to a sound recording.

Vocabularies

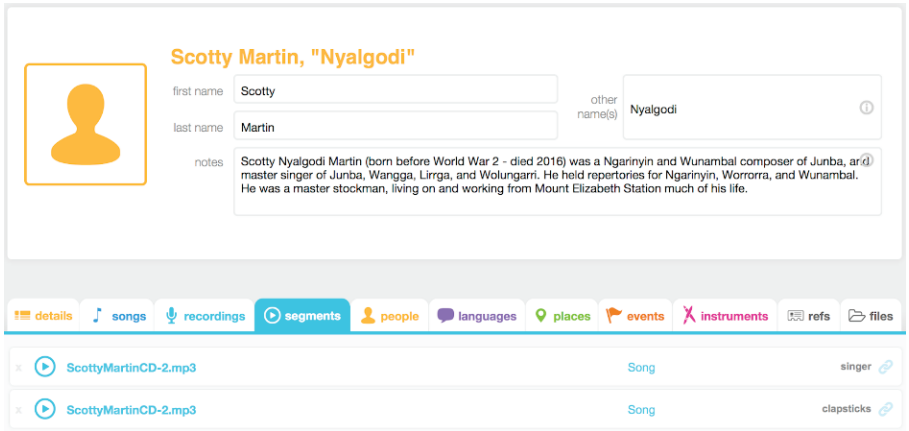
A controlled vocabulary is a selected list of terms used by cataloguers to consistently describe resources in a collection. The CCCT uses several unstandardised vocabularies, outlined below, in order to accurately describe First Nations music material. These vocabularies provide additional context to the metadata, such as the instruments featured in a piece of recorded material. However, the terms found within each CCCT vocabulary cannot be perfectly mapped to another single vocabulary, as will be realised below.

[33] 'RDF 1.1 Concepts and Abstract Syntax,' World Wide Web Consortium (W3C), 25 February 2014, https://www.w3.org/TR/rdf11-concepts.

Roles

The CCCT includes a controlled vocabulary of roles played by people involved in a recording. The roles contained in the CCCT vocabulary include those often attached to performers, creators and researchers involved in First Nations song. Several of these roles can be crosswalked to the OLAC standard, while others cannot, bringing into question the suitability of OLAC. It is also worth noting the OLAC roles vocabulary has not been updated since 2006. A more suitable alternative is the more comprehensive Library of Congress Relators vocabulary, as will be demonstrated below.[34] The roles in the CCCT database include recordist, singer, lead singer, clapstick player, didjeridu player, rasp player, guitarist, speaker, other participants, transcriber, annotator and rights holder.

Figure 6. A CCCT record linking Scotty 'Nyalgodi' Martin to a recording segment with the roles of singer and clapstick player.



In the case where appropriate URLs or other unique identifiers cannot be found among the Library of Congress Relators vocabulary, string literals can be used as a last resort. That is to say, the phrase 'guitarist' can be used instead of a URL. However, this limits the linking functionality of the metadata and leaves it open to anomalies (for example, typographical errors or varied capitalisation). This is a particular area where further research will be required.

Table 3. A comparison of terms available in the OLAC and Library of Congress Relators vocabularies.

Current term	OLAC equivalent	Library of Congress Relators Equivalent
Speaker	http://www.language-archives.org/REC/role.html#speaker	http://id.loc.gov/vocabulary/relators/spk
Transcriber	www.language-archives.org/REC/role.html#transcriber	http://id.loc.gov/vocabulary/relators/trc
Producer	None	http://id.loc.gov/vocabulary/relators/pro
Sound engineer	None	http://id.loc.gov/vocabulary/relators/sde
Guitarist	None	None

[34] 'Relators,' Library of Congress, accessed 10 June 2024, https://id.loc.gov/vocabulary/relators.html.

Instruments

Several thesauri exist that can provide unique identifiers (URLs) for the instruments featured in the CCCT database. These include the Library of Congress Medium of Performance Thesaurus for Music (LCMPTM), MusicBrainz instrument list, and AIATSIS Topical Thesaurus (formerly the AIATSIS Subject Thesaurus).[35] Several instruments feature in recordings stored in the CCCT, including the didjeridu, body percussion, clapsticks, rasp and guitar. The didjeridu, body percussion, clapsticks and guitar have entries in both the LCMPTM and MusicBrainz instrument list, however the rasp does not feature in either. The AIATSIS Topical Thesaurus, meanwhile, contains entries for all five instruments, and will be the proposed solution for this purpose. A case could be made to use the LCMPTM because it is constructed with linked data functionality in mind, while the AIATSIS Topical Thesaurus is not. This decision should be reviewed later to determine whether the AIATSIS Topical Thesaurus is the most appropriate option.

Geographic Locations

Similar to instruments, several thesauri exist to provide identifiers for geographical areas. These include the Getty Thesaurus of Geographic Names (TGN), the GeoNames database and AIATSIS Place Thesaurus.[36] Many of these options are suitable for our purposes, to varying degrees, as they offer linked data functionality (except AIATSIS) and a persistent identifier to link and enrich our metadata. After consideration, the TGN is recommended for this crosswalk, as it is purpose-built for linked data and has extensive coverage of the Australian continent.

One limitation that must be addressed is that places stored in the CCCT are likely to be First Nations ceremonial grounds, remote communities, or other locations that may not have an equivalent entry in the TGN. Given how important the connection is between place and performance, this limitation is important to note as it current hinders efforts to future-proof First Nations music metadata. A workaround to this limitation would be to assign unique identifiers to locations featured within the collection material, and subsequently link them to other geographic regions listed in the TGN to give more context. One example of this is the Bijili ceremony ground, which can be given an ad-hoc identifier in its own right and then linked using the Schema.org property `containedInPlace` to the TGN identifier for the Kimberley region (`7032377`) (see Fig. 7).

First Nations language and ownership

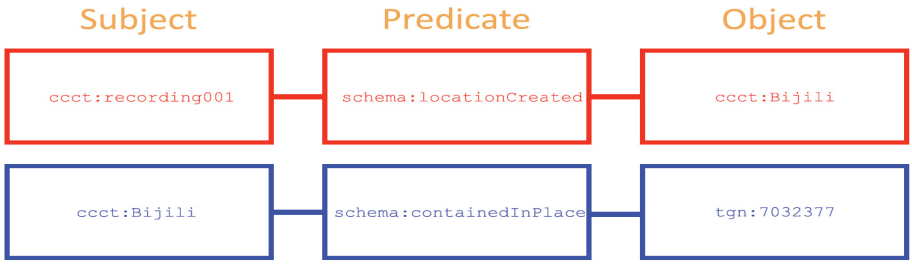
In addition to recording relevant geographic metadata, the CCCT database also has a relation for the various languages included in songs and recordings. This is a fairly simple crosswalk to execute, as it can draw from the AIATSIS Austlang thesaurus. The Austlang thesaurus contains

---

[35] AIATIS, 'Topical Thesaurus,' accessed 17 November 2024, https://aiatsis.gov.au/thesaurus/topical; Library of Congress, 'Library of Congress Medium of Performance Thesaurus for Music,' accessed 9 June 2024, https://id.loc.gov/authorities/performanceMediums.html; 'Instrument List,' MusicBrainz, accessed 9 June 2024, https://musicbrainz.org/instruments.

[36] AIATSIS, 'Place Thesaurus,' accessed 17 November 2024, https://aiatsis.gov.au/thesaurus/place; 'GeoNames,' Geonames, accessed 10 June 2024, https://www.geonames.org; 'Getty Thesaurus of Geographic Names,' Getty Research Institute, accessed 10 June 2024, https://www.getty.edu/research/tools/vocabularies/tgn.

Figure 7. A diagram demonstrating the links between a CCCT recording, the Bijili ceremonial ground, and the Getty Thesaurus entry for the Kimberley.



URLs for each First Nations language group and references the group's alternate names or spellings, such as Noongar and Nyoongar.[37]

One outstanding issue is finding an appropriate equivalent for the 'firstNation' field in the Places relation. The purpose of the 'firstNation' field is to record the traditional owner(s) of a geographic location, however this specific kind of ownership is not represented in Schema.org. Possible Schema.org equivalents include `owns` or `landlord`, but these are intended for commercial products and real estate, respectively, and are therefore insufficient. This issue speaks to Toner's and Corn and Patrick's points about the limitations of Western frameworks when applied to non- Western knowledge.[38] This is an opportunity for future research.

Repertories, Segments and Collections

Many of the resources recorded in the CCCT will be contained within other resources. For instance, a recording can comprise multiple independent segments, several songs can make up a repertory, and a collection can include multiple recordings. In most cases, the Schema.org property `isPartOf`, or its inverse `hasPart`, can be used to link a child resource to its parent. As it stands, the CCCT does not have separate relations (or tables) for repertories. Additionally, there is no Repertory class in any of the ontologies used in this project. There is, however, a Schema.org class for `Collection`, and this could be used for both the repertories and collections identified in the CCCT, though it risks losing specificity.

Events

Within the CCCT's scope, events can include performances, recording sessions and non-physical occurrences such as the conception of a musical work during a dream. All of these can be categorised under the Schema.org class `Event`. The CCCT's Events relation is fairly simple, with very few fields, all of which can be easily crosswalked.

Notes and Comments

Several of the CCCT relations contain fields for notes, often produced by researchers or collection managers. The notes can relate specifically to other fields such as the date, source, or research undertaken. These notes serve a specific purpose in the CCCT and an equivalent

[37] 'W41: NOONGAR / NYOONGAR,' AIATSIS Collection, accessed 8 June 2024, https://collection.aiatsis.gov.au/austlang/language/w41.

[38] Toner, 'History, Memory and Music,' 14; Corn and Patrick, 'Exploring the Applicability of the Semantic Web,' 133.

Schema.org class will not be available for each type of note. This may be an occasion when some specificity is lost during the crosswalk. Instead, it is proposed that all notes are converted to the class `schema:Comment`, where multiple comments can be assigned to a resource such as a recording or song.

Conclusion

The purpose of the project was to crosswalk the metadata schema of the CCCT database to a more standardised format. The rationale behind this was to ensure that metadata stored in a CCCT instance becomes more interoperable with other systems, thereby improving accessibility to the material. To do this, the project involved proposing a set of ontologies, thesauri and other standards that can act as appropriate linked data equivalents. It sought to do this within the greater context of current accessibility and interoperability issues that databases containing First Nations music metadata currently experience.

Crosswalking data stored in a relational database, such as the CCCT, to a non-relational format, such as a linked data knowledge graph, represents a radical change in how this data is structured. However, this project should be seen as an attempt to 'future-proof' the database for the eventual adoption of linked data more broadly by libraries and archives. Linked data's adoption is already underway, as evidenced by the Library of Congress and other major institutions. Though uptake has been slow, there has been a noticeable embrace of linked data locally by related organisations such as PARADISEC and the Language Data Commons of Australia (LDACA).

As mentioned, several thesauri maintained by AIATSIS were considered or adopted as part of this project. Though entries in these thesauri have a persistent URL, they are not structured for linked data purposes, which is not ideal for making as many connections as possible through a knowledge graph. For instance, broader terms are not linked to narrower ones through subject-predicate-object triples. Converting the AIATSIS thesauri into linked data ontologies—and updating them to account for new entries in the process—could result in a valuable and unique resource for describing First Nations cultural material that is currently unmatched by any other ontology or thesaurus.

Carrying out the CCCT crosswalk project involved acknowledging several unresolved issues, mainly around the lack of standards that directly accommodate the complexities of metadata associated with First Nations music material. The limits of Western metadata standards to accurately describe the complexities of First Nations cultural material have previously been highlighted, and it is clear that current metadata frameworks remain insufficient.[39] An ideal standard would account for the various relationships present within First Nations communities, such as moieties, connections to place, and specific access protocols (restricted between men, restricted between women, etc.). Some work has already commenced on these efforts, though more can still be done.

About the Author

Caleb Triscari is a graduate researcher at the University of Melbourne and a member of the *Context* editorial committee. His research focuses on musicology / ethnomusicology, music preservation and sustainability, archival studies, and music information retrieval.

[39] Toner, 'History, Memory and Music;' Corn and Patrick, 'Exploring the Applicability of the Semantic Web.'