BIBLIOTHEK – Forschung und Praxis 2018; 42(2): 206–221 **DE GRUYTER**

Terhi Nurmikko-Fuller*, Daniel Bangert, Alan Dix, David Weigl and Kevin Page

Building Prototypes Aggregating Musicological Datasets on the Semantic Web

https://doi.org/10.1515/bfp-2018-0025

Abstract: Semantic Web technologies such as RDF, OWL, and SPARQL can be successfully used to bridge complementary musicological information. In this paper, we describe, compare, and evaluate the datasets and workflows used to create two such aggregator projects: *In Collaboration with In Concert*, and *JazzCats*, both of which bring together a cluster of smaller projects containing concert and performance metadata.

Keywords: Musicology; ontology; workflow

Die Erstellung prototypischer Anwendungen von verknüpften musikwissenschaftlichen Datensätzen

Zusammenfassung: Semantische Web-Technologien wie RDF, OWL und SPARQL ermöglichen die Verknüpfung von komplementären musikwissenschaftlichen Daten. In diesem Artikel beschreiben, vergleichen und bewerten wir die Datensätze und Workflows, die zur Erstellung zweier solcher Aggregationsprojekte verwendet wurden: *In Collaboration with In Concert* und *JazzCats*, die jeweils Sammlungen kleinerer Projekte mit Konzert- und Performance-Metadaten zusammenführen.

Schlüsselwörter: Musikwissenschaft; Ontologie; Workflow

Inhalt

***Kontaktperson: Dr. Terhi Nurmikko-Fuller,** terhi.nurmikko-fuller@anu.edu.au
Dr. Daniel Bangert, bangert@sub.uni-goettingen.de
Prof. Alan Dix, alan@hcibook.com
Dr. David Weigl, david.weigl@oerc.ox.ac.uk
Dr. Kevin Page, kevin.page@oerc.ox.ac.uk

1 Introduction

Diverse research agendas in the area of digital musicology result in the production of complementary but often disconnected data capturing information about musical works, composers, and performers in their wider historical and cultural contexts. The combination of existent traditional research paradigms, the tacit knowledge of domain-experts, and the affordances of the increasingly semantic Web enable the discovery of musicological information in a new, rich data environment. The interlinking of datasets that have been published in machine-processable formats such as RDF[1] and the use of Semantic Web technologies (e.g. Linked Data,[2]

1 The RDF acronym refers to the Resource Description Framework model. It is a World Wide Web Consortium (W3C) standard for publishing information online in a machine-processable and interchangeable way. For further information see https://www.w3.org/RDF/.

2 Linked Data is a publication paradigm, which utilises existing Web architecture and technologies to bring about a Web of Data (cf. the current manifestation of the World Wide Web as a Web of Documents). HTTP URIs point to specific instances of data, and the relationship between them (rather than to webpages). If the information represented in this way is accessible to human users and software agents freely and without restrictions, we consider it to be Linked Open Data. For more information about Linked Data see https://www.w3.org/standards/semanticweb/data.

RDF,[3] and SPARQL[4]) enable new digital methods for scholarly investigation. Such bridging of data presents challenges to expert musicologists and data scientists when working with legacy tabular or relational datasets that do not natively facilitate linking and referencing to and from external sources. The problems of reconciliation brought on by different schemas, data types, and limited instance-level[5] overlap have been tackled through the creation of an interconnected knowledge graph of linked RDF triples,[6] in which information can be retrieved and discovered. Here, we present a number of pragmatic approaches for turning legacy datasets into RDF, and outline the heuristics applicable to each described workflow. Both aggregator projects contain relational databases and tabular data, and the process of data conversion is neither automatic nor, given the musicological considerations of the data, straightforward. The production of RDF that adequately captures the knowledge contained within all the sub-projects requires domain expertise and, simultaneously, the use of existing tools requires familiarity with them and their limitations. Description of the heuristics and evaluation of the final workflow are essential.

Extant Linked Data projects (such as Pelagios project[7] or Europeana[8]) have illustrated the use of instance-level and class-level (type-based) alignments between datasets. Although the capture of workflows is not unprecedented,[9] few research projects have actively sought to reapply documented workflows in an effort to prove reusability. It is this assessment of the reproducibility of workflows that has influenced and inspired the repetition of the *InC-InC* (In Collaboration with In Concert) workflow in the context of *JazzCats* (Jazz Collection of Aggregated Triples).[10]

We begin with an introduction to Linked Data in general (Section 2), carrying on to provide an overview of existing work in the field of digital musicology (Section 3). In Section 4, we describe two projects that integrate related datasets about music performance. These projects make use of five datasets in total and each contain information about musical performances, associated ephemera, and applicable metadata. Section 5 illustrates the ontological structures used as part of the RDF production workflows, which themselves are outlined in Section 6. The penultimate section (7) provides an evaluation of these structures and workflows through comparative analysis between the two aggregator projects, and a view to future work.[11]

2 A brief introduction to Linked Data

The Semantic Web is a vision and set of technologies to enable machine-readable data to be shared on the Web as easily as (web) pages allow the sharing of human-readable text.[12] Standard relational database systems such as MySQL and MS Access can export and import data tables using CSV files, describe the contents of a table using a database schema, and query the data using SQL. The Semantic Web

3 The RDF data model is used to represent information. It enables data exchange, even between systems with different underlying organisational schemas.

4 SPARQL is a recursive acronym (the SPARQL Protocol And RDF Query Language). As the name suggests, it is a tool for querying and manipulating data expressed as RDF and held in a graph database (or triplestore). For further information see https://www.w3.org/2009/sparql/wiki/Main_Page.

5 The task of making explicit to a machine that which is implicit to a human is completed, in part, through the division of information into data categories. In the context of Semantic Web technologies, information structures known as ontologies (further elaborated on in footnote 15 and 37) are used to capture the general patterns of data types contained in the dataset (such as people or places) and the relationships between them. These are schema-level representations of a domain. Instance-level data entities refer to the specific individuals that populate these data categories (such as Roy Eldridge, or Berlin).

6 RDF is expressed through clusters of HTTP URIs, most often in sets of three (hence, triples), referred to as the subject, the predicate, and the object. The predicate represents the relationship that connects the subject and the object. For further information see https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.

7 http://commons.pelagios.org/.

8 https://www.europeana.eu/.

9 Bechhofer et al. (2013a), Missier et al. (2010).

10 Nurmikko-Fuller et al. (2016).

11 The work for *Transforming Musicology* and its subprojects was supported by the UK Arts and Humanities Research Council (AH/L006820/1) through the *Digital Transformations* theme. The work reported on here has been supported by the EPSRC *Fusing Semantic and Audio Technologies for Intelligent Music Production and Consumption (FAST IMPACt)* project (EP/L019981/1). The authors also want to thank their colleagues David Lewis, John Pybus, and Carolin Rindfleisch, University of Oxford; Richard Lewis, Tim Crawford, and Simon McVeigh, University of London Goldsmiths; Rachel Cowgill, University of Huddersfield; Rupert Ridgewell, British Library; and Christina Bashford, University of Illinois at Urbana-Champaign. We also want to thank Professor José Antonio Bowen, Goucher College, for the Body and Soul data; Martin Pfleiderer, Hochschule für Musik Franz Liszt Weimar, and the Jazzomat Research Project team for the Weimar Jazz Database; and M. Cristina Patuelli, Pratt Institute, and the Linked Jazz team for their dataset; as well as Stefan Münnich, Universität Basel, for *JazzCats* user-testing. We extend a special thank you to our collaborator Dr Alfie Abdul-Rahman, University of Oxford, as the creator of the *JazzCats* project website, and one of the original members of the *JazzCats* team.

12 Berners-Lee et al. (2001).

has corresponding technologies to those above and used on the document Web:[13] RDF for data interchange, OWL[14] for describing data ontologies[15] and SPARQL for querying.

These newer technologies and formats better support the explicit capture of meaning (semantics). In an Excel worksheet, the user knows that the 'price' column will contain amounts of money, or the 'employee' table in a database will describe a person; the meaning is in the heads of those using the data. For automatic web sharing, data may be picked up from anywhere, so a way of determining meaning needs to be explicitly encoded in the data: RDF and OWL add precisely this level of semantics. For example, if representations of concerts exist in two different datasets, they can be coded to explicitly refer to the same type of event even if the datasets were produced by entirely different teams of people.

When accessing a web page, users can follow links to discover more information about things. Linked Data enables analogous behaviours on the Semantic Web.[16] Linked Data employs Uniform Resource Identifiers (URIs) to identify data records or metadata entries. Instead of using local database identifiers such as 'AH37' to refer to a concert, a dereferenceable URI is used.[17] The contents retrieved from this URI provide machine-readable data about the concert. This approach aids discoverability: the user doesn't need to know about the location of data before starting and can simply follow links from dataset to dataset.

3 Related Work

The application of Semantic Web technologies to provide aggregated access to interlinked musical information has been previously proposed by specialist communities within musicology.[18] They have been successfully applied in the context of *Transforming Musicology*,[19] *SALAMI: Structural Analysis of Large Amounts of Music Information*,[20] and the Répertoire International de Littérature Musicale.[21] RISM: Répertoire International des Sources Musicales[22] is a further example of a similar research agenda. These projects have resulted in publications[23] and workshops such as Digital Libraries for Musicology, co-located with the Joint Conference on Digital Libraries in 2014[24] and 2015,[25] and the International Society of Music Information Retrieval annual conference in 2016[26] and 2017[27]. Linked Data has also been applied to performance studies,[28] crowdsourced musicological recommendations,[29] live music archives,[30] and concert programme ephemera, as will be described below. Semantic Web techniques such as ontologies and reasoning have also been used to build a working set of Linked Data.[31] Ontological developments currently under way within the larger context of digital musicology include structures mapping the nature of leitmotifs,[32] as well as an extension or revision[33] of the CHARM[34] ontology.

In the work described here, we made use of a number of existing ontologies: FOAF (Friend of a Friend ontology, for describing people, their activities, and interpersonal relationships),[35] and SKOS (Simple Knowledge Organisation System, a standard for representing thesauri, taxonomies, and other classification schemes in the context of the Semantic Web),[36] the more domain-specific Music,[37] Event,[38] and Timeline[39] ontologies, as well as Schema.org (used to describe structured data on webpages),[40] and the bibliographic metadata ontologies of Bibframe[41] and FaBiO.[42] Although widely used, the existing ontologies outlined here were insufficient to completely map all avail-

13 As described earlier (footnote 2).
14 The Web Ontology Language (OWL) is computational logic-based Semantic Web language. OWL documents are known as ontologies. For more information https://www.w3.org/OWL/.
15 Ontologies are OWL documents, used to represent and define the concepts and internal relationships inherent within a dataset or domain in a machine-processable format. For more information on ontologies, please see https://www.w3.org/standards/semanticweb/ontology.
16 Heath and Bizer (2011).
17 http://example.org/c/AH37.
18 De Roure (2014), De Roure et al. (2015).
19 Crawford et al. (2014).
20 Bay et al. (2009).
21 http://www.rilm.org/.
22 https://opac.rism.info/metaopac/start.do?View=rism.
23 Bashford et al. (2000).
24 http://dl.acm.org/citation.cfm?id=2660168.
25 http://dl.acm.org/citation.cfm?id=2785527.
26 http://dl.acm.org/citation.cfm?id=2970044.
27 https://dl.acm.org/citation.cfm?id=3144749
28 Page et al. (2015).
29 Musto et al. (2013), Adamou et al. (2014).
30 Bechhofer et al. (2013b), Page et al. (2017).
31 Dix et al. (2010).
32 Dreyfus and Rindfleisch (2014).
33 Harley and Wiggins (2015).
34 Wiggins and Harris (1990).
35 Brickley and Miller (2014).
36 Miles et al. (2005).
37 Raimond and Giasson (2007).
38 Raimond and Abdallah (2007).
39 Raimond and Abdallah (2006).
40 http://schema.org/docs/schemas.html.
41 https://www.loc.gov/bibframe/.
42 Shotton and Peroni (2011).

able data. As a result, some new ontological development formed part of the workflow for the projects presented here (see Section 4).

Disambiguation between entities in the datasets was achieved with the use of existing external Linked Data authority URIs, namely VIAF,[43] DBpedia,[44] MusicBrainz,[45] Wikidata,[46] and the BBC.[47]

4 Describing the data

We describe the data, ontological models, and workflows used to convert five separate datasets into RDF. These data represent the content of two distinct projects comprising information regarding music performances and their associated ephemera and metadata. These aggregator projects are *InC-InC* and *JazzCats*. Both contain data produced in their own distinct sub-projects.

While there are some instance-level parallels and matches between the datasets of these aggregator projects, it is rather data structure similarities that enabled us to validate the reproducibility of our workflows.[48] Specifically, both aggregator projects include at least one sub-project containing only tabular data and at least one other sub-project where information is held in a relational database. Table 1 contains a representative sample illustrating the similarities between datasets, as well as the unique features of their data.

Table 1: Representative Sample of Data Categories across all sub-projects

Aggregator projects	In Concert		JazzCats		
Data category \ Subprojects	LC18	LC19	Body&Soul	WJazzD	Linked Jazz
Place	✓	✓	✓		
Title	✓	✓	✓	✓	
Performance Type	✓			✓	
Event Metadata	✓		✓		
Performance		✓	✓		
Ephemera	✓	✓			
Person		✓	✓	✓	✓
Musical Work	✓	✓	✓	✓	
Instrument			✓	✓	
Digital Signal Metadata				✓	

43 https://viaf.org/.
44 http://wiki.dbpedia.org/
45 https://musicbrainz.org/
46 https://www.wikidata.org/wiki/Wikidata:Main_Page.
47 Raimond et al. (2010).
48 Nurmikko-Fuller et al. (2016), Nurmikko-Fuller et al. (2017).

4.1 In Collaboration with In Concert

In Collaboration with In Concert (InC-InC)[49] was a small-scale investigation into the workflow necessary to enable the publication of musicological data on the Web in a machine-processable format (namely RDF). Recorded and published earlier,[50] this workflow was repeated for *JazzCats* (section 4.2).[51] Before we describe the developed workflow and subsequent *InC-InC* project, *In Concert: Towards a Collaborative Digital Archive of Musical Ephemera* (InConcert) warrants description and discussion.

4.1.1 In Concert: Towards a Collaborative Digital Archive of Musical Ephemera

InConcert is a collaborative project examining performance metadata (collected from concert ephemera, such as programmes, bills, reviews, adverts, and other information) sourced from historical newspapers and periodicals, as well the bibliographical metadata of those primary sources.[52] It was undertaken within the larger *Transforming Musicology* project,[53] funded by the UK Arts and the Humanities Research Council,[54] which ran between 2013 and 2017. InConcert contains data from three separate sub-projects: *Calendar of London Concerts 1750-1800* (LC18),[55] *19th-century London Concert Life (1815-1895)* (LC19),[56] and OCR (Optical Character Recognition) derived data from the British Musical Biography (BMB).[57] The aim of InConcert was to

49 Nurmikko-Fuller et al. (2016).
50 Nurmikko-Fuller et al. (2016).
51 http://jazzcats.oerc.ox.ac.uk/.
52 Nurmikko-Fuller et al. (2016).
53 http://transforming-musicology.org/about/.
54 http://www.ahrc.ac.uk/.
55 McVeigh (n.y.).
56 Bashford (2003).
57 https://archive.org/details/britishmusicalb00brow, Brown and Stratton (1897).

create a musicological digital library[58] that would connect the LC18 and LC19 datasets, to enable trends and patterns to be examined across over 150 years of concerts in London.

4.1.1.1 Calendar of London Concerts 1750–1800 (LC18)

Calendar of London Concerts 1750–1800 (LC18) data and associated documentation are openly available as tabular data (Creative Commons Attribution NonCommercial ShareAlike CSV and XLS).[59] Based on a stable dump of the LC18 database, these CSV files were transformed to JSON and imported into a noSQL database. Many of the data categories contain information which is accessible to human users using a cross-referencing system with available documentation, but are inaccessible to software agents: much of the information is captured in acronyms, for example 'CNS' for 'Casino, Great Marlborough Street' (the performance venue), or 'GB' for 'Garden Benefit' (event type). The ontological modelling carried out as part of the *InC-InC* workflow[60] sought to capture this implicit information and represent it explicitly in a machine-processable format.

4.1.1.2 19th-century London Concert Life

19th-century London Concert Life (1815–1895) (LC19) is comprised of bibliographical metadata regarding concert ephemera: data instances refer to pamphlets, newspapers, and other historical print material which contain information and details about performances, including their locations and artists involved. Based on a legacy Oracle database dump, the data is contained within a MySQL database, with a structure more complex than that of the tabular LC18 outlined above. Instance-level data for LC19 is not publicly shared, but was made available to the research team for the *InC-InC* workflow.[61]

4.2 JazzCats (Jazz Collection of Aggregated Triples)

JazzCats (Jazz Collection of Aggregated Triples)[62] was originally conceived as a Semantic Web project, hosted within Virtuoso,[63] a well-established open-source triplestore that manages RDF data. The project combines three previously distinct datasets into one Virtuoso instance and enables them to be queried from a single entry point.[64] This unified knowledge base is further interlinked with data in external sources (VIAF, DBpedia, MusicBrainz, Wikidata, and the BBC), and enables scholars to ask new kinds of research questions about jazz performance history and the social and professional relationships between musicians.

As an aggregator project, *JazzCats* amalgamates data from three different sub-projects: the *Body and Soul* discography (Body&Soul); the *Weimar Jazz Database* (WJazzD), which contains metadata about jazz solo performances such as instrument, style, duration, tempo, and key; and a previously established Linked Data project that publishes the social and professional relationships between jazz musicians, Linked Jazz.

4.2.1 Body and Soul discography

Body and Soul discography (Body&Soul) describes over 200 recordings of the jazz standard *Body and Soul*, all made between 1930 and 2004. This discography was originally published as a supplement to *Who plays the tune in "Body and Soul"? A performance history using recorded sources*.[65] This information is available as a PDF file from the author's website,[66] but this data publication method is representative of only 'one star' Linked Open Data;[67] that is, it is available on the web, and has an open licence, but is not represented in a machine-readable form. It was therefore not directly included in the workflow for this project: rather, a CSV file provided by the author through personal correspondence, and enriched prior to conversion to RDF (see Section 6). The data cleaning and enriching process was carried out in OpenRefine[68] and included the clustering and normalization of performer names, instruments, and dates. The resulting dataset derived from the original CSV file is openly available (Creative Commons Attribution NonCommercial).[69]

58 Bainbridge et al. (2014).
59 http://datatodata.com/in-concert/LC18/list.php?type=concerts.
60 Nurmikko-Fuller et al. (2016).
61 Nurmikko-Fuller et al. (2016).
62 http://jazzcats.oerc.ox.ac.uk/.
63 http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main.
64 http://jazzcats.oerc.ox.ac.uk/sparql.
65 Bowen (2015).
66 http://josebowen.com/body-and-soul/.
67 https://www.w3.org/DesignIssues/LinkedData.html.
68 http://openrefine.org/.
69 Bangert (2016).

4.2.2 Weimar Jazz Database (WJazzD)

Weimar Jazz Database (WJazzD)[70] is an extensively curated and verified collection of transcriptions of jazz solo performances (covering a range of artists and various subgenres) from the Jazzomat Research Project.[71] Although copyright restrictions prevent access to note and contextual annotations, temporal markers associated with MusicBrainz IDs make the identification of existing solos possible.[72] The data contain specifics regarding the performers, instruments, and titles of musical works, as well as musicological metadata such as style, tempo, key, and other features of the digital signals for each recording. WJazzD links to external authority files for artists (Wikipedia URIs) and recordings (MusicBrainz URIs).

4.2.3 Linked Jazz

Linked Jazz[73] is a pre-established RDF resource capturing a prosopography of jazz musicians, queryable from a single access point.[74] The project focus lies in capturing the social and professional relationships between musicians, ranging from rel:friendOf[75] to mo:collaboratedWith[76] and the Linked Jazz project-specific lj:inBandTogether,[77] as well as several other gradients on the socio-professional scale. Disambiguation within the dataset is achieved through linking to external authorities such as the Library of Congress (LoC)[78] and DBpedia.[79]

5 Ontology design and knowledge representations

In order to successfully complete the data format conversion from tabular or relational data structures into a knowledge graph, each of the datasets described in Section 4.2 were mapped onto a bespoke ontological structure by a musicologist with additional expertise in data librarianship. With the exception of the model used for Body&Soul (described in Section 5.3), classes and properties from existing ontologies and schemas were used in conjunction with project-specific ones. Each of these structures is described in detail below.

5.1 Ontology for LC18

For LC18, the research team created a new TTL[80] file with a bespoke ontological structure that contained classes and properties from existing ontologies (see Fig. 1). While both the LC18 and Body&Soul ontological structures relied extensively on existing classes and properties from the Music Ontology,[81] RDFS,[82] OWL, SKOS,[83] Geo,[84] and Event,[85] the former also incorporates bibliographical metadata ontologies; namely Bibframe, FaBiO,[86] and Schema.org. Project-specific properties were defined for InC:is_performance_type, InC:venue_for, InC:reviewed_in, InC:listed_in, InC:prog_for, InC:advertises, InC:is_advertised_in, InC:has_title, and InC:has_ticket. Classes were created for InC:Performance_Type, InC:Programme, InC:Advert, InC:Title, and InC:Price. At the heart of the model are entities which are equally mapped as instances of both mo:Performance and event:Event.

5.2 Ontology for LC19

Data for LC19 was captured as RDF through a largely automated workflow (see Section 6.2). This resulted in both the knowledge-graph structure and the instance level data being mapped onto the generic vocab: namespace. SPARQL queries were used to modify the resulting graph to provide mappings to the FOAF, Schema.org, and Bibframe ontologies, with additional project-specific properties asserted for InC:occupation (for employment status of a person), and InC:captured_in_record, which connects a per-

70 http://jazzomat.hfm-weimar.de/dbformat/dboverview.html.
71 http://jazzomat.hfm-weimar.de/dbformat/dbcontent.html.
72 Abeßer et al. (2014).
73 https://linkedjazz.org.
74 https://linkedjazz.org/sparql/.
75 http://vocab.org/relationship/#knowsOf.
76 http://motools.sourceforge.net/doc/musicontology.html#term_collaborated_with.
77 The full <URI> for this property is https://linkedjazz.org/ontology/inBandTogether but that does not, unlike the other examples in this paper, point to documentation.
78 http://id.loc.gov/authorities/names.html.
79 http://wiki.dbpedia.org/.
80 TTL (pronounced as "Turtle" and referring to the Terse RDF Triple Language), is a syntax for RDF. It has similarity to SPARQL and can be read by human users with relative ease. It is also considered to be easier to manually edit than alternatives such as RDF/XML. For more information on TTL, please see https://www.w3.org/TR/turtle/.
81 Raimond and Giasson (2007).
82 https://www.w3.org/TR/rdf-schema/.
83 https://www.w3.org/2004/02/skos/.
84 http://www.w3.org/2003/01/geo/.
85 Raimond and Abdallah (2007).
86 http://www.sparontologies.net/ontologies/fabio/source.html.

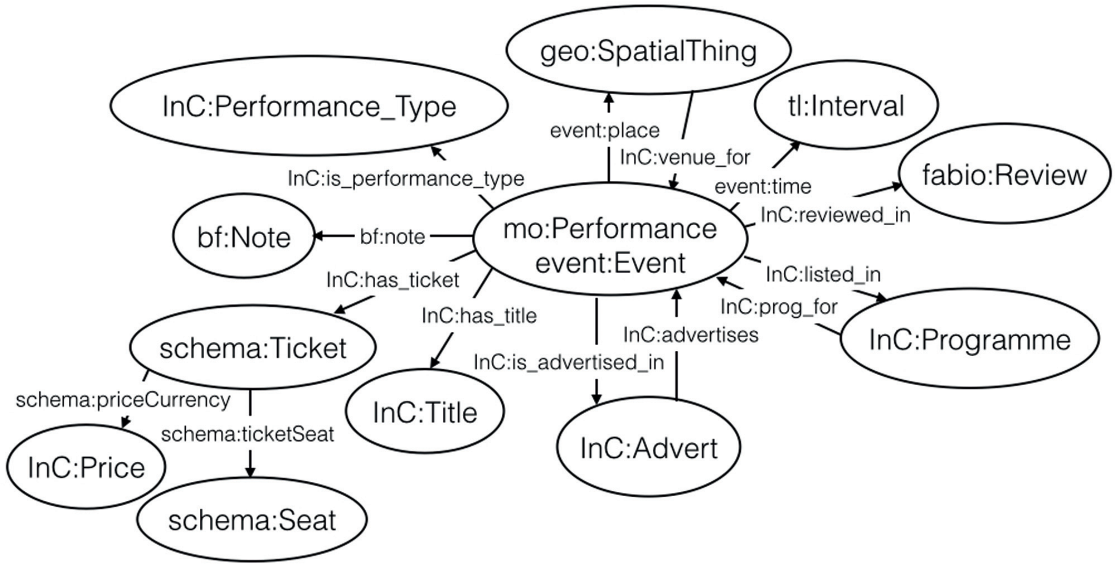

Fig. 1: Ontological structure for LC18

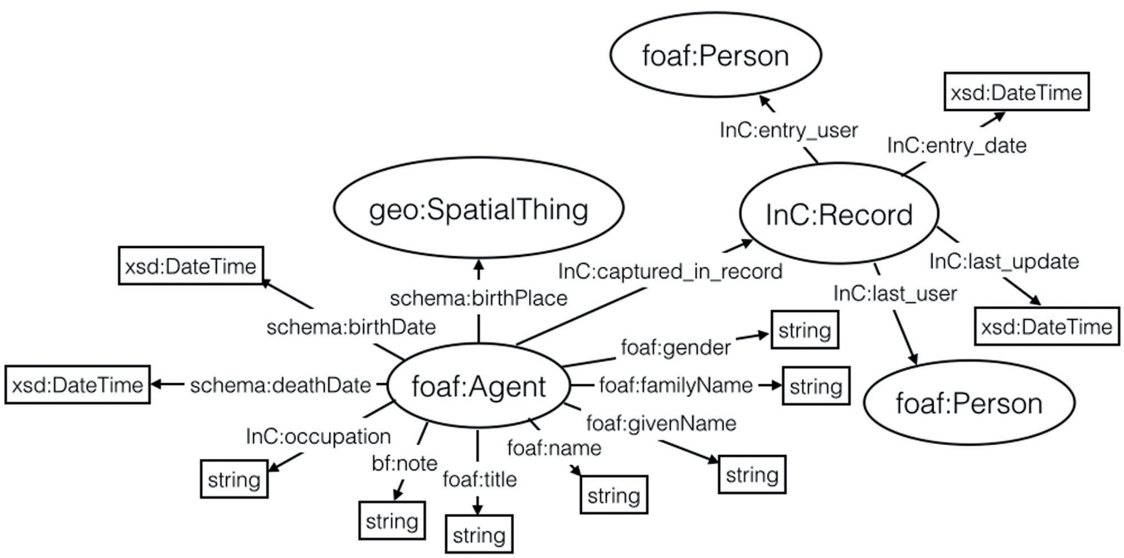

Fig. 2: Person section of the LC19 ontological structure

son who appears in the content of a metadata record to the appropriate record. This enabled us to assert a specific creation date, and a most recent update for a metadata record, as well as describe users who accessed the metadata record as separate types of person from those who appear in the content of the metadata record. This separation of the metadata record and the person described in the content of the ephemera is captured in Fig. 2.

5.3 Ontology for Body&Soul

For Body&Soul, existing ontologies were imported from the Web directly, using URIs, with classes and properties selected according to the model illustrated in Fig. 3. In comparison to LC18's ontological structure, Body&Soul was mapped much more extensively to the classes and properties of the Music Ontology. Equivalence is expressed using skos:closeMatch based on the need to link concepts that may not always be completely interchangeable.[87] Although

87 Halpin et al. (2010).

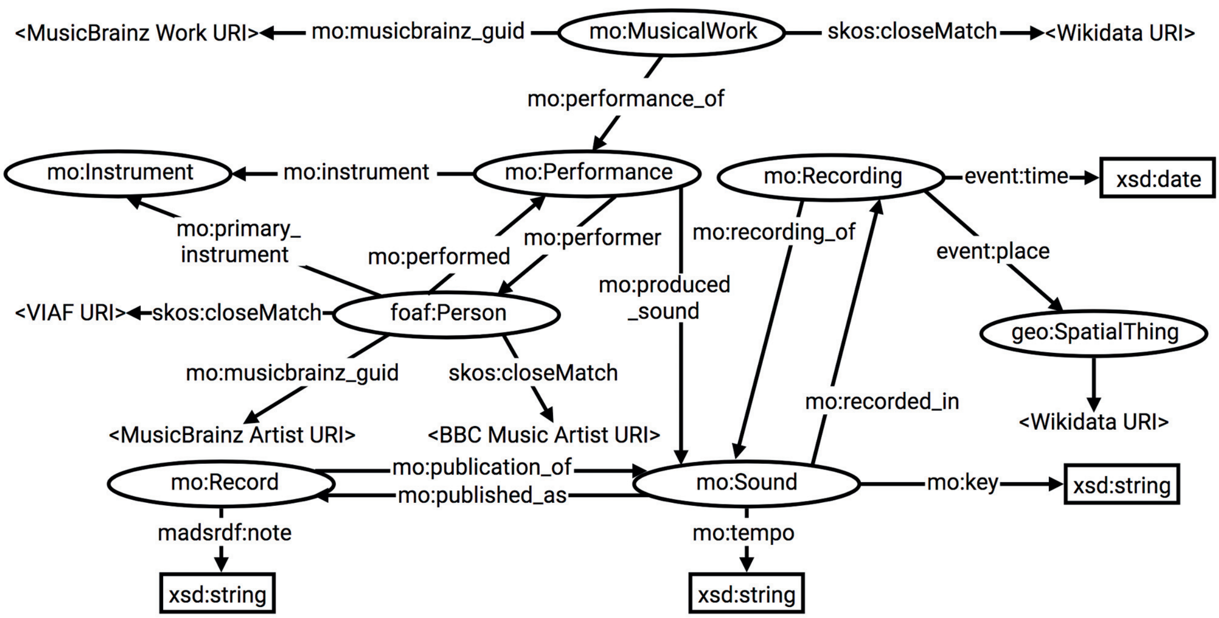

Fig. 3: Ontological structure for Body&Soul

other datasets in the *JazzCats* project required project-specific properties and classes to be used, none were necessary for the representation of the Body&Soul data.

5.4 Ontology for WJazzD

The workflow (described in Section 6.2) used for the production of RDF triples representing the information contained within WJazzD was a largely automated one, reproducing the steps outlined for the data conversion for LC19.

The WJazzD ontological structure stands out from the others in the *JazzCats* aggregator project (see Table 1) as preliminary analysis of the data yielded relatively few opportunities for mapping to existing ontologies or schemas. As a result, the majority of the classes and properties used (and illustrated in Fig. 4a) are project-specific in the jazzcats: namespace.

The structure of the WJazzD database was faithfully captured in the resulting RDF triples, which, with little reinterpretation or change result in the centralised graph structures depicted in Fig. 4a and Fig. 4b. To avoid confusion arising from similar information category types,[88] the illustrations of the ontological structure capture the different URI schema used for the data sections (see Section 6.4). Future iterations of the project will examine whether a simpler or a less centralised graph could be used to streamline the model into a more effective and computationally efficient structure.

The dataset also contains many instances where xsd: string and xsd:integer were used to capture the value of the property (see Fig. 4b). For textual or numerical properties such as jcm:duration, jcm:beatdur, and the various WJazzD internal IDs this is unproblematic, since the value of the property has no inherent semantics. There are, however, several opportunities for further semantic enrichment. These include the representation of the values described by properties such as jcsi:rhythmfeel, mo:key, and jcsi:style in musicological meaningful information categories.

5.5 Ontology for Linked Jazz

Linked Jazz is the third sub-project within *JazzCats*. It is a pre-established Linked Data project, with RDF triples available for download from the project website.[89] These data are based around a simple ontological model with only one class (foaf:Person[90]) and some 30 different properties; a mix of established (e.g. foaf:name,[91] foaf:depic-

88 Clusters of properties as depicted in Fig. 4b for jcv:solo_info and jc:Melody occur for each of the other data types (classes) depicted in Fig. 4a, namely jcv:composition_info, jcv:melody_info, jcv:record_info, jcv:sections, jcv:tack_info, and jcv:beats.

89 https://linkedjazz.org/access/.

90 http://xmlns.com/foaf/spec/#term_Person.

91 http://xmlns.com/foaf/spec/#term_name.

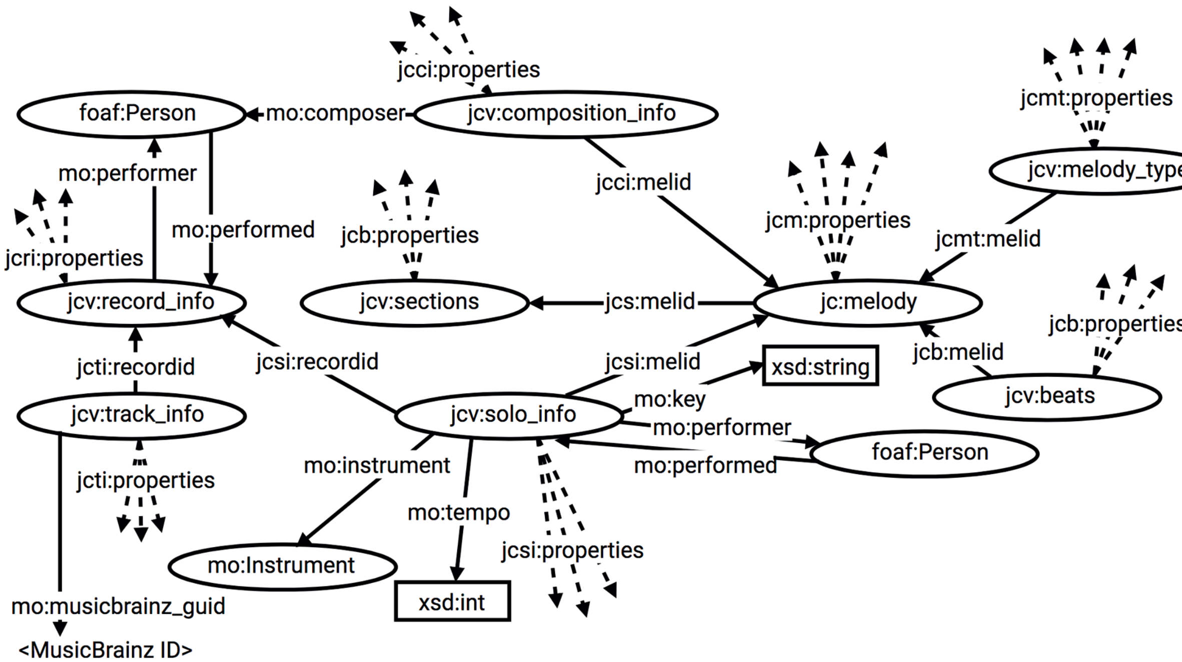

Fig. 4a: An Ontological structure of the overall WJazzD dataset

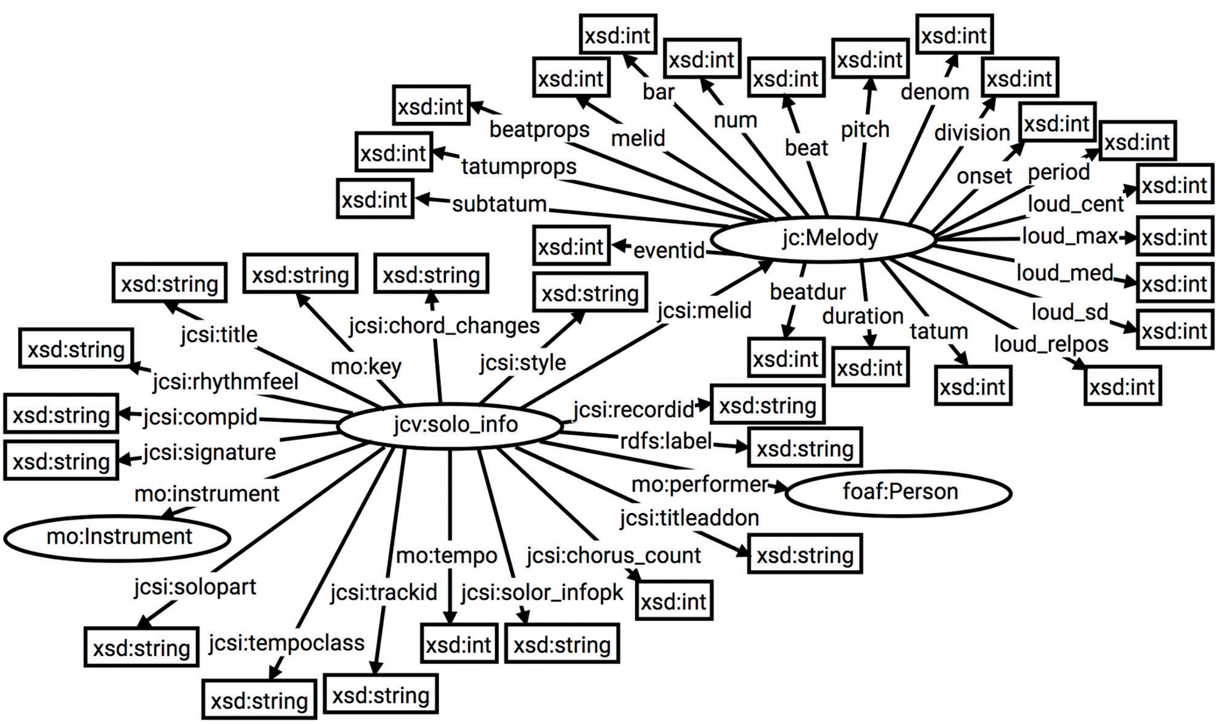

Fig. 4b: Detail from the WJazzD ontological structure

tion[92]) and project-specific properties (e.g. lj:playedTogether, lj:touredWith, and lj:bandLeaderOf).

This dataset was ingested into *JazzCats* as existing RDF triples, and no design decisions regarding the underlying ontological modelling were carried out. The appearance of foaf:Person in the ontology visualised in Fig. 5 reflects our decision to incorporate a legacy dataset (see Section 6.3). This also prompted us to define people in the

92 http://xmlns.com/foaf/spec/#term_depiction.

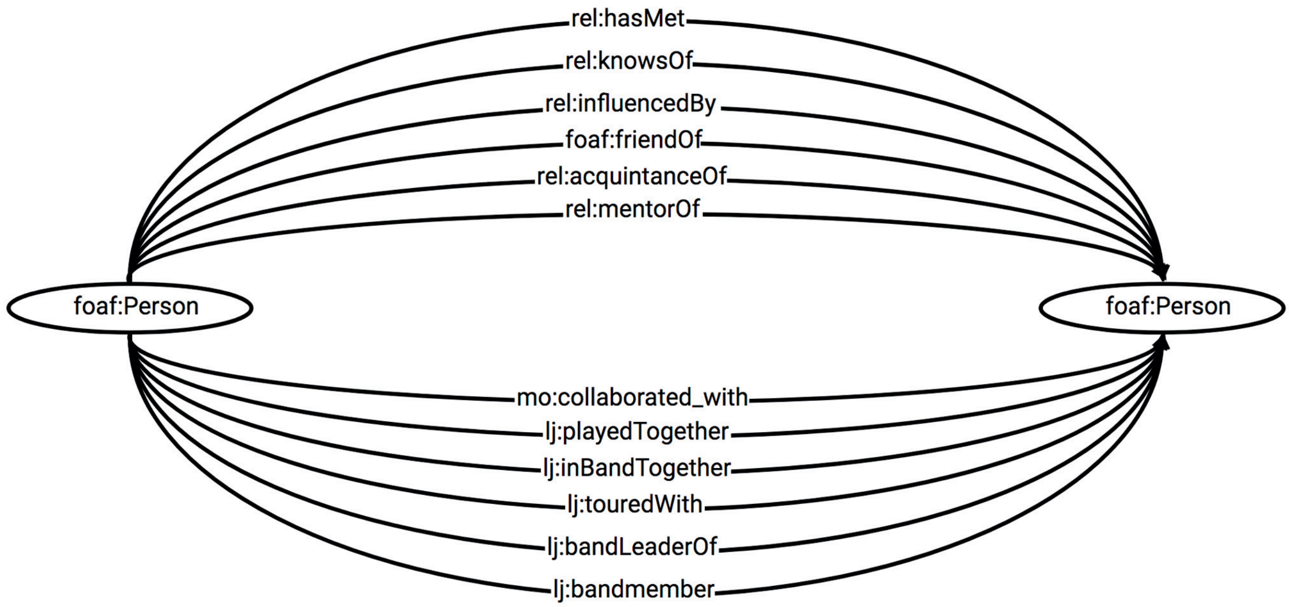

Fig. 5: Ontological structure for Linked Jazz

other datasets using the same class, so as to enable schema-level alignment between all the *JazzCats* sub-projects.

6 Methodology and workflow

Semantic Web technologies, when applied not only to the capture of instance-level data, but also the underlying information structures and workflows used to produce them, have the potential to allow the bridging of disparate but complementary datasets in digital musicology.[93] This can be particularly useful when collaborative projects bring together the diverse data, methods, and foci of several researchers. The similarities between the data types, information structures, and necessary workflows for RDF production of the aggregator projects *InC-InC* and *JazzCats* have provided an opportunity to evaluate the reproducibility of the methods applied to the former in the context of the latter.

6.1 Workflow for producing RDF using Web-Karma

Both *InConcert* and *JazzCats* contain tabular data. For *InConcert*, this is the LC18 dataset, described in Section 4.1.1.1. For *JazzCats*, it is Body&Soul (Section 4.2.1). These two datasets contain similar types of performance metadata (people, places, etc.), but it is the data structures of these sets which enable the repetition of an identical workflow.

The data from both LC18 and Body&Soul was converted to RDF using open-source software called Web-Karma,[94] produced by the University of Southern California and made available for download and use.[95] The software has some dependencies (Apache Maven 3.0[96] and Java 1.7[97]). Once Web-Karma has been installed, the user must upload both the data, and either upload or import RDF files containing relevant ontologies. This involves deciding which ontological structures to upload and use (for example, if they have designed and produced their own), or whether to import one or more existing ontologies. Whilst Web-Karma accepts other syntaxes (e.g. RDF/XML), the best user experience is achieved when using the more human-readable TTL. Upon successful uploading, Web-Karma will recognise the TTL file as an OWL ontology. The steps for uploading are then repeated for the dataset. The Web-Karma UI can be used in a point-and-click process to assign semantic value to each category of data. Assigning an appropriate value is simplest when using a CSV file, which is shown as separate columns for each data type (or class). Web-Karma's functionality in-

93 Nurmikko-Fuller and Page (2016).

94 http://usc-isi-i2.github.io/karma/.
95 https://github.com/usc-isi-i2/Web-Karma/wiki.
96 https://maven.apache.org/docs/3.0/release-notes.html.
97 http://www.oracle.com/technetwork/java/javase/downloads/index.html.

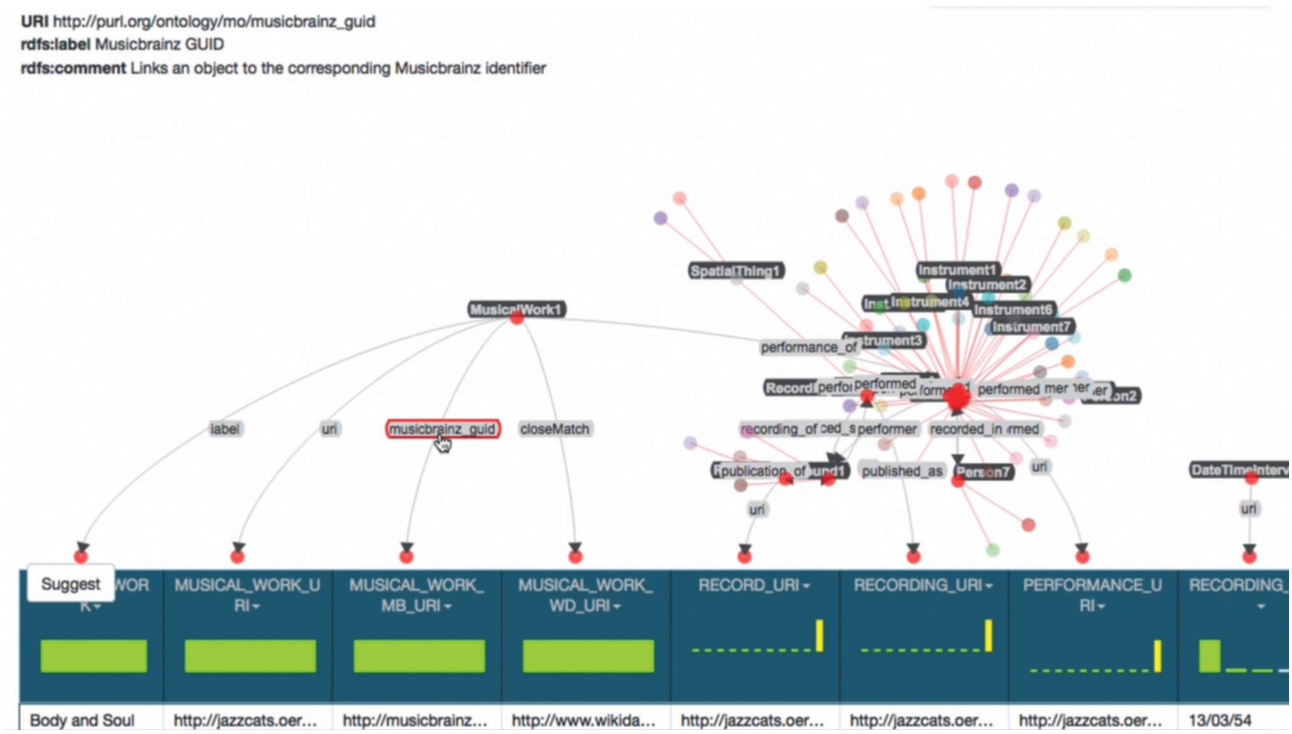

Fig. 6: Web-Karma user interface

cludes visual representation of the resulting knowledge graph (fig. 6).

The limitation of this software is the lack of up-to-date and clear documentation capturing the semantic value assignment (i.e. the alignment of the ontological class to a given column of data). Before mapping tabular data to a specified ontology, the user must have a very clear understanding of both the data and the ontological structure. Reviewing the ontology is not possible in the user interface (UI), although mapped entity types and their connecting relationships are visualised in a dynamic graph (see Fig. 6). The ambiguity of the labels within the UI (for example, referring to the individuals that populate a class as being “Properties of a Class”) means that the process of assigning semantic values can appear more complex than it is.

The benefit of using this tool is that the resulting RDF should require minimal post-hoc editing if produced by an expert with a clear understanding of the ontological model and familiarity with the data. In the case of Body&Soul, manual edits were only required for a small number of URIs which had been minted based on entity labels, and contained some syntactical errors (such as spaces and commas).

6.2 Workflow for producing RDF using D2RQ

InC-InC and *JazzCats* both contain sub-projects where data is held in a relational database; for the former, LC19 data held in MySQL; for the latter, WJazzD data stored in SQLite3. Both databases made it possible to carry out a largely automated workflow using a pre-existing open-source tool, D2RQ.[98] Although a largely automated process, running D2RQ against a relational database requires two iterations of this stage of the workflow (Fig. 7): the first, to capture the database structure, and the second to populate the knowledge graph with instance-level data.

The resulting RDF was, in both LC19 (part of *InC-InC*) and WJazzD (in *JazzCats*), batch-edited using SPARQL queries. A conscious decision was made to make every effort to map the elements of both datasets to existing ontologies (Sections 5.2 and 5.4). Preference was given to solutions that mirrored those applied to the other projects: people were represented using FOAF; musicological features were captured using relevant classes and properties in the Music Ontology. For LC19, most of the ontological structure relies on existing properties and classes. For WJazzD, the vast majority of the properties and classes are project-specific, since for many of the data types and their

98 http://d2rq.org/ and http://d2rq.org/d2r-server.

relationships, no existing ontologies containing appropriate classes and properties were identified.

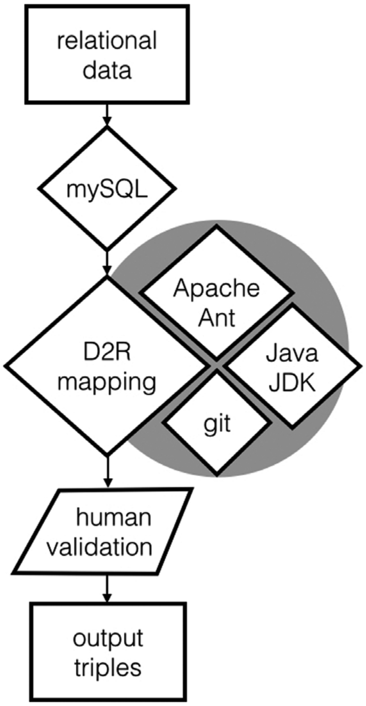

Fig. 7: Workflow for using D2RQ with the LC19 data in a MySQL database

One noticeable difference between the two datasets was an additional step in the WJazzD workflow, introduced by the absence of primary keys within the SQLite3 database. The issue was solved by running commands over the relational tables inside SQLite3 to add primary keys where necessary. Command line tools (generate_mapping, dump–rdf) were used to generate TTL capturing the database structure and to generate instance-level RDF triples respectively.

This approach is well-suited to the task of producing RDF from large, structurally complex databases, which could not have been mapped within the technical parameters of Web-Karma (see Section 6.1). The challenges of using this tool are largely related to the insufficiently documented stages of the initial install and setup of D2RQ, and the steps necessary to align the application with the database. The RDF triples produced using this method also require later edits to more accurately align them with the appropriate ontological structure, since the ones produced in this automated process capture the structure of the database. For example, running D2RQ on the WJazzD data, the relationship between a specific solo performance and the instrument was captured, but needed to be edited using SPARQL queries to be mo:instrument.

An additional step following the Web-Karma and D2RQ workflows for *InC-InC* was to add an RDF data plugin to the InConcert data API. This enables users to access these data as RDF alongside the previously available JSON and CSV formats.

6.3 Workflow for ingesting existing RDF (Linked Jazz)

For datasets already published as RDF, data can be ingested to a local triplestore or queried remotely if an endpoint is available. For example, in the case of Linked Jazz, access to published RDF is provided via a SPARQL endpoint.[99] When considering how to include Linked Jazz data in *JazzCats*, remote querying was tested and several issues were encountered.[100] The decision was then made to ingest three Linked Jazz data-dumps (people,[101] relationships,[102] and a name directory[103]) into the *JazzCats* triplestore. The authors recognize the possible need to re-ingest whenever changes or updates are introduced to the Linked Jazz triples.

Some issues were encountered during the addition of Linked Jazz RDF into the *JazzCats* triplestore. Correcting them resulted in a deviation from the original data dump, and thus a deviation of the triples available from the Linked Jazz website. These changes were:

– An error in the URI for Martin Luther King Jr., found in RDF representing people (Jr. <http://xmlns.com/foaf/0.1/name> "Martin Luther King"@en). The string "Jr." was changed to the DBpedia URI (http://dbpedia.org/resource/Martin_Luther_King,_Jr.).
– People are not defined as instances of a class such as foaf:Person as might be expected.[104] As a result, the RDF could only be linked to the other projects' data at instance-level, rather than entity type. To solve the problem, we added an earlier Linked Jazz dataset (the Linked Jazz Name Directory),[105] which contains class attributions, to our triplestore.
– There was some ambiguity regarding individuals contained within the dataset. This is illustrated by the rdfs:comment associated with both http://linkedjazz.

99 https://linkedjazz.org/sparql/.
100 For example, the Linked Jazz SPARQL endpoint not appearing to filter results when DISTINCT was included as part of a query.
101 http://linkedjazz.org/api/people/all/nt.
102 http://linkedjazz.org/api/relationships/all/nt.
103 https://linkedjazz.org/data/jazz_directory_aug_2012.nt.
104 Pattuelli et al. (2015).
105 https://linkedjazz.org/data/jazz_directory_aug_2012.nt.

org/resource/Ed_Jobear and http://linkedjazz.org/resource/Hal_Serra.[106]

Where the datasets contained valid RDF, they were left unaltered. For a small number of occurrences of broken triples in the Linked Jazz data-dumps, the appropriate DBpedia URI was corrected prior to ingestion into the project triplestore. The authors recognize this as a deviation from the original data, and as a step that may have to be repeated in the future, as and when new versions of the Linked Jazz triples are added to *JazzCats*. To facilitate and enable the repeatability of the ingest and transformation process, these changes have been documented and are publicly available through the *JazzCats* website.[107]

7 Evaluation and discussion

Working in an interdisciplinary team of musicologists, ontologists and information engineers involves collaborative decision-making balancing musicological concerns with the affordances of Semantic Web technologies. As prototypes, *InC-InC* and *JazzCats* demonstrate a robust and repeatable process of data modelling and integration, and the potential to leverage a diverse set of skills in pursuit of musicological research questions.

7.1 Design decisions

Domain expertise was used to validate data enrichment and integration at several stages of the *InC-InC* and *JazzCats* projects. In *JazzCats*, this was done directly by a musicologist[108] and both projects involved collaborative ontology design to create knowledge graphs that can be accurately navigated. To illustrate this process in greater detail, we outline how musicological aims guided processes of organising and validating data for InConcert.

Early work on InConcert identified a number of key musicological concerns for the project and indeed digital archives in general. These included the desire to be: *authoritative* and of known quality, so that the data can be used reliably for further interpretation, and *complete*, or at least sampled in a well-controlled and well-documented manner, so that bias in any trends observed or statistical analysis derived from the data is minimised.[109]

This led to two design decisions: first, the project did not adopt the common practice of drawing multiple datasets into a single combined dataset with the ability to re-extract the separate datasets as views if needed. While this would have made combining the data easy it would have the potential to hide differences in collection methodology and interpretation that led to the datasets.

The amalgamated data of InConcert could be suitably tagged to retain provenance and allow specific musicologists the ability to update their own parts of the combined dataset. However, this form of access-related ownership does not at present elicit the same confidence as clearly separate files or databases, even though these may themselves share the same underlying storage disks.

The original datasets of InConcert come in formats that are familiar to the musicologists and have existing archival practices and third-party use. If amalgamating the datasets had led to the need for new update mechanisms and different ways of accessing the data, it would have broken those existing practices.

Hence the data organisation of InConcert retains the original documents and datasets as the 'golden copy' and uses a form of federated access to provide the data in a common external form including user querying, and a JSON and CSV data API. This does include some caching of the source data, some additional data to encode links between datasets, and meta-descriptions of individual data tables and collections to allow the different datasets to be viewed in a relatively consistent manner. However, the overall access mechanisms follow the "the leaves are golden" information design principle[110] retaining the original data as far as possible.

The second design decision was to ensure that when there was any level of automated data enhancement, this was clearly marked in the datasets and subject to expert validation. One example of this was entity (or instance-level) reconciliation between the datasets, matching venues and people. Expert validation by musicologists was performed using a combination of bespoke interfaces and downloadable spreadsheets that could be edited and re-uploaded.[111] Common to all was that the intelligent matching algorithms employed in these interfaces were liberal in selecting potential matches, but that these were always shown to the musicologists to verify and much more con-

106 The comment reads: ""He is a dentist, can't find a website for him."@en".
107 http://jazzcats.oerc.ox.ac.uk/documentation/.
108 Bangert (2016).
109 Quoting Dix et al. (2014).
110 Dix (2016).
111 Dix et al. (2016).

servative measures used to highlight those that are potentially problematic.

7.2 Enabled research questions

By structuring, aggregating and publishing datasets as Linked Open Data, *InC-InC* and *JazzCats* enable music scholars to construct queries that draw on previously unconnected information. For instance, *JazzCat*s allows musicological analysis to shift between discographic information, performance features (style, tempo, key), and the professional and social networks of an artist. Research questions that are enabled by *JazzCats*[112] include:

- Which performances of *Body and Soul* were recorded in a particular style in a specific place? For example, swing performances recorded in London.
- Which recordings of *Body and Soul* feature a particular combination of instruments, in a specific key? For example, recordings with trumpet and piano, performed in the key of D-flat.
- Which performances of *Body and Soul* were recorded in a specific place by artists that played with a particular artist? For example, identify recordings of *Body and Soul* made in New York City by artists who played with Roy Eldridge during their career.
- What is the relationship between artists that recorded *Body and Soul*? For example, the relationships between artists connected to trumpet player Roy Eldridge.

The enabled research questions demonstrate how *JazzCats* can assist to contextualize and contest work on jazz performance histories.

7.3 Future work

The current manifestation of *JazzCats* is of a functioning prototype. Future development will see the ingestion and addition of additional discographic sources, such as J-DISC,[113] which is an example of session-based data that could provide valuable additional information about recordings and professional networks if published as Linked Data.[114] Other work will include improving the internal connectivity by disambiguating between identifiers, and aligning instances referring to the same musicians, performances, and recordings.[115]

Although the *InC-InC* and *JazzCats* projects have made data available as RDF Linked Data, they effectively represent two mostly separate islands of data with few interchanges. They each act individually as exemplars of interlinking within their own 'island' of data and this is valuable in itself, but, as yet, they are a first tentative step towards fully demonstrating the potential for Linked Open Data. They do, however, show what might be possible in future.

Consider Wigmore Hall, a London concert hall built in 1901. Despite lying just outside the coverage date of LC19, a selection of early 20th century concerts at Wigmore Hall was used as an early demonstrator of LC19.[116] Wigmore Hall holds considerable paper archives and aims to digitise them; when this is completed, they will connect well into the InConcert datasets. Whilst still retaining a classical repertoire, Wigmore Hall now also hosts a Jazz series, and so starts to interconnect with *JazzCats*. It is clear that, as more datasets are added to the Linked Data web of musicological data, the current isolated data islands will join and allow rich analysis across periods and genres.

8 Conclusion

The discussed workflows highlight methodological options and challenges involved in structuring and publishing of Linked Data on the Web. The enabled queries demonstrate how access to semantically integrated data can assist scholars to document, analyse, and interpret music-related event data as captured in performance ephemera and recordings. The complete and comprehensive capture of all information within the projects described here remains an avenue of further development and research. For both aggregator projects, the inclusion of symbolic and audio data with the existing metadata would improve the range of educational and scholarly use cases. In terms of user experience and accessibility, further methods of querying, visualising and analysing these data could assists scholars

112 SPARQL queries for JazzCats data can be found at http://jazzcats.oerc.ox.ac.uk/access/ and https://github.com/terhinurmikko/JazzCats.

113 http://jdisc.columbia.edu/.

114 Hao et al. (2016).

115 For example, linking recordings within *JazzCats* to MusicBrainz Recording URIs will enrich the project knowledge graph. Information about performers and instruments on specific recordings is currently only partially available due to the lack of complete information about secondary performers in Body&Soul (labeled 'Other Performers' in the original dataset).

116 http://inconcert.datatodata.com/Wigmore-Hall.

wanting to take full advantage of potential research applications.

Bibliography

Abeßer, Jakob; Cano, Estefanía; Frieler, Klaus; Pfleiderer, Martin (2014): Dynamics in jazz improvisation – Score-informed estimation and contextual analysis of tone intensities in trumpet and saxophone solos. In: *Proceedings of the 9th Confer. Interdisciplinary Musicology (CIM),* Berlin, Germany, 156–61.

Adamou, Alessandro; d'Aquin, Mathieu; Barlow, Helen; Brown, Simon (2014): LED: Curated and crowdsourced linked data on music listening experiences. In: *Proceedings ISWC 2014 Posters & Demonstrations Track within the 13th Int. Semantic Web Conference, ISWC 2014*, Riva del Garda, Italy, 93–96.

Bainbridge, David; Hu, Xiao; Downie, J. Stephen (2014): A Musical Progression with Greenstone: How Music Content Analysis and Linked Data is Helping Redefine the Boundaries to a Music Digital Library. In: *Proceedings of the ACM 1st International Workshop on Digital Libraries for Musicology*, New York, NY, USA, 1–8.

Bangert, Daniel (2016): JazzCats Body and Soul discography [dataset]. Zenodo. Available at http://doi.org/10.5281/zenodo.163886.

Bashford, Christina; Cowgill, Rachel; McVeigh, Simon (2003): The Concert Life in Nineteenth-Century London Database Project. In: *Nineteenth-Century British Music Studies*. Aldershot: Ashgate, 1–12.

Bay, Mert; Burgoyne, John Ashley; Crawford, Tim; De Roure, David; Downie, J. Stephen; Ehmann, Andreas; Fields, Ben; Fujinaga, Ichiro; Page, Kevin; Smith, Jordan B.L. (2009): Structural Analysis of Large Amounts of Music Information.

Bechhofer, Sean; Ainsworth, John; Bhagat, Jiten; Buchan, Iain; Couch, Philip; Cruickshank, Don; De Roure, David; Delderfield, Mark; Dunlop, Ian; Gamble, Matthew; Goble, Carole; Michaelides, Danius; Missier, Paolo; Owen, Stuart; Newman, David; Sufi, Shoaib (2013a): Why Linked Data is not enough for scientists. In: *Generation Computer Systems,* 29(2), 599–611.

Bechhofer, Sean; Page, Kevin; De Roure, David (2013b): Hello cleveland! Linked data publication of live music archives. In: *14th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2013*, Paris, France, 1–4.

Berners-Lee, Tim; Hendler, James; Lassila, Ora (2001): The Semantic Web. In: *Scientific American*, May, 29–37.

Bowen, José Antonio (2015): Who plays the tune in "body and soul"? A performance history using recorded sources. In: *Journal of the Society of American Music,* 9(3), 259–92.

Brickley, Dan; Miller, Lilly (2014): FOAF Vocabulary Specification 0.99. Namespace Document 14 January 2014. Paddington Edition.

Brown, James; Stratton, Stephen Samuel (1897): British musical biography: a dictionary of musical artists, authors, and composers born in Britain and its colonies. SS Stratton.

Crawford, Tim: Fields, Ben; Lewis, David; Page, Kevin (2014): Explorations in Linked Data practice for early music corpora. In: *Digital Libraries, IEEE*, 309–12.

De Roure, David (2014): Executable Music Documents. In: *Proceedings of the ACM 1st International Workshop on Digital Libraries for Musicology*, 1–3.

De Roure, David; Klyne, Graham; Page, Kevin; Pybus, John; Weigl, David (2015): Music and Science: Parallels in Production. In: *Proceedings of the ACM 2nd International Workshop on Digital Libraries for Musicology*, 17–20.

Dix, Alan (2016): The Leaves are Golden – putting the periphery at the centre of information design. Keynote at HCI2016, July 2016, Bournemouth, UK.

Dix, Alan; Cowgill, Rachel; Bashford, Christina; McVeigh, Simon; Ridgewell, Rupert (2014): Authority and Judgement in the Digital Archive. In*: the ACM 1st International Digital Libraries for Musicology workshop, ACM/IEEE Digital Libraries conference 2014*, London 12th Sept, 1–8.

Dix, Alan; Cowgill, Racher; Bashford, Christina; McVeigh, Simon; Ridgewell, Rupert (2016): Spreadsheets as User Interfaces. In: *Proceedings ACM AVI 2016*, 192–95.

Dix, Alan; Katifori, Akrivi; Lepouras, Giorgos; Vassilakis, Costas; Shabir, Nadeem (2010): Spreading activation over ontology-based resources: From personal context to web scale reasoning. In: *International Journal of Semantic Computing, Special Issue on Web Scale Reasoning: scalable, tolerant and dynamic*, 4(1), 59–102.

Dreyfus, Laurence; Rindfleisch, Carolin (2014): Using Digital Libraries in the Research of the Reception and Interpretation of Richard Wagner's Leitmotifs. In: *Proceedings of the ACM 1st International Workshop on Digital Libraries for Musicology*, 1–3.

Halpin, Harry; Hayes, Patrick J.; McCusker, James P.; McGuinness, Deborah L.; Thompson, Henry S. (2010): When owl:sameas isn't the same: An analysis of identity in Linked Data. In: *International Semantic Web Conference*, 305–20.

Hao, Yun; Choi, Kahyun; Downie, J. Stephen (2016): Exploring J-DISC: Some Preliminary Analyses. In: *Proceedings of the 3rd International workshop on Digital Libraries for Musicology*, 41–44.

Harley, Nicholas; Wiggins, Geraint (2015): An Ontology for Abstract, Hierarchical Music Representation. In: *Proceedings of 16th International Society for Music Information Retrieval Conference.*

Heath, Tom; Bizer, Christian (2011): Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web. In: *Theory and Technology, 1(1), 1–136.*

McVeigh, Simon: Calendar of London Concerts 1750–1800. Goldsmiths, Dataset. University of London. Available from http://research.gold.ac.uk/10342/.

Miles, Alistair; Matthews, Brian; Wilson, Michael, and Brickley, Dan (2005): SKOS Core: Simple Knowledge Organisation for the Web. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 3–10.

Missier, Paolo; Sahoo, Satya; Zhao, Jun; Goble, Carole; Sheth, Amit: Janus (2010): from workflows to semantic provenance and Linked Open Data. In: McGuinness D.L., Michaelis J.R., Moreau L. (eds) Provenance and Annotation of Data and Processes. IPAW 2010. *Lecture Notes in Computer Science*, vol 6378. Springer, Berlin, Heidelberg.

Musto, Cataldo; Narducci, Fedelucio; Semeraro, Giovanni; Lops, Pasquale; de Gemmis, Marco (2013): Distributional models vs. linked data: Exploiting crowdsourcing to personalize music playlists. In*: Proceedings 4th Italian Information Retrieval Workshop*, Pisa, Italy, 84–87.

Nurmikko-Fuller, Terhi; Bangert, Daniel; Abdul-Rahman, Alfie (2017): All the Things You Are: Accessing An Enriched Musicological

Prosopography Through JazzCats. In: *Digital Humanities 2017 Conference, Montreal,* Canada, August 8–11.
Nurmikko-Fuller, Terhi; Dix, Alan; Weigl, David; Page, Kevin (2016): 2016, August. In Collaboration with In Concert: Reflecting a Digital Library as Linked Data for Performance Ephemera. In: *Proceedings of the ACM 3rd International workshop on Digital Libraries for Musicology*, 17–24.
Nurmikko-Fuller, Terhi; Page, Kevin (2016): A linked research network that is Transforming Musicology. In: *Proceedings of the 1st Workshop on Humanities in the Semantic Web co-located with 13th ESWC Conference* 2016 (ESWC 2016), Anissaras, Greece, May 29th, 73–78.
Page, Kevin; Bechhofer, Sean, Fazekas, Gyorgy; Weigl, David; Wilmering, Thomas (2017): Realising a Layered Digital Library: Exploration and Analysis of the Live Music Archive through Linked Data. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 1–10.
Page, Kevin; Nurmikko-Fuller, Terhi; Rindfleisch, Carolin; Weigl, David; Lewis, Richard; Dreyfus, Laurence; De Roure, David (2015): A toolkit for live annotation of opera performance: Experiences capturing Wagner's ring cycle. In: *Proceedings 16th International Conference of Music Information Retrieval*, 411–17.
Pattuelli, Cristina; Provo, Alexandra; Thorsen, Hilary (2015): Ontology building for linked open data: A pragmatic perspective. In: Journal of Library Metadata, 15(3-4), 265–94.
Raimond, Yves; Abdallah, Samer (2006): The Timeline Ontology. Available at http://motools.sourceforge.net/timeline/timeline.html.
Raimond, Yves; Abdallah, Samer (2007): 2007. The Event Ontology. Technical report. Available at http://motools.sourceforge.net/event/event.html.
Raimond, Yves; Giasson, Frédérick (2007): Music Ontology Specification. Available at http://motools.sourceforge.net/doc/musicontology.html.
Raimond, Yves; Scott, Tom; Sinclair, Patrick; Miller, Libby; Betts, Stephen; McNamara, Frances (2010): Case study: Use of semantic web technologies on the BBC web sites. In: *W3C Semantic Web Use Cases and Case Studies*. Available at https://www.w3.org/2001/sw/sweo/public/UseCases/BBC/.
Shotton, David; Peroni, Silvio (2011): FaBiO: FRBR Aligned Bibliographic Ontology. Available at http://www.sparontologies.net/ontologies/fabio/source.html.
Wiggins, Geraint; Harris, Mitch; Smaill, Alan (1990): Representing music for analysis and composition. University of Edinburgh, Department of Artificial Intelligence.

Dr. Terhi Nurmikko-Fuller
Centre for Digital Humanities Research
Australian National University
120 McCoy Circuit
ACT 2601
Australia
terhi.nurmikko-fuller@anu.edu.au

Dr. Daniel Bangert
Universitätsbibliothek Göttingen
Universität Göttingen
Platz der Göttinger Sieben 1
D-37073 Göttingen
bangert@sub.uni-goettingen.de

Prof. Alan Dix
HCI Centre
School of Computer Science
University of Birmingham
UK
alan@hcibook.com

Dr. David Weigl
Oxford e-Research Centre
Department of Engineering Science
University of Oxford
7 Keble Road
Oxford OX1 3QG
UK
david.weigl@oerc.ox.ac.uk

Dr. Kevin Page
Oxford e-Research Centre
Department of Engineering Science
University of Oxford
7 Keble Road
Oxford OX1 3QG
UK
kevin.page@oerc.ox.ac.uk