Graduate Student Syncopation:
Contributing to LINKED JAZZ

Pratt SCHOOL OF INFORMATION & LIBRARY SCIENCE
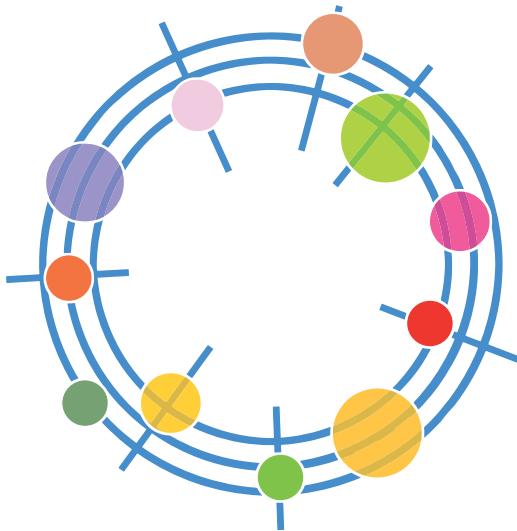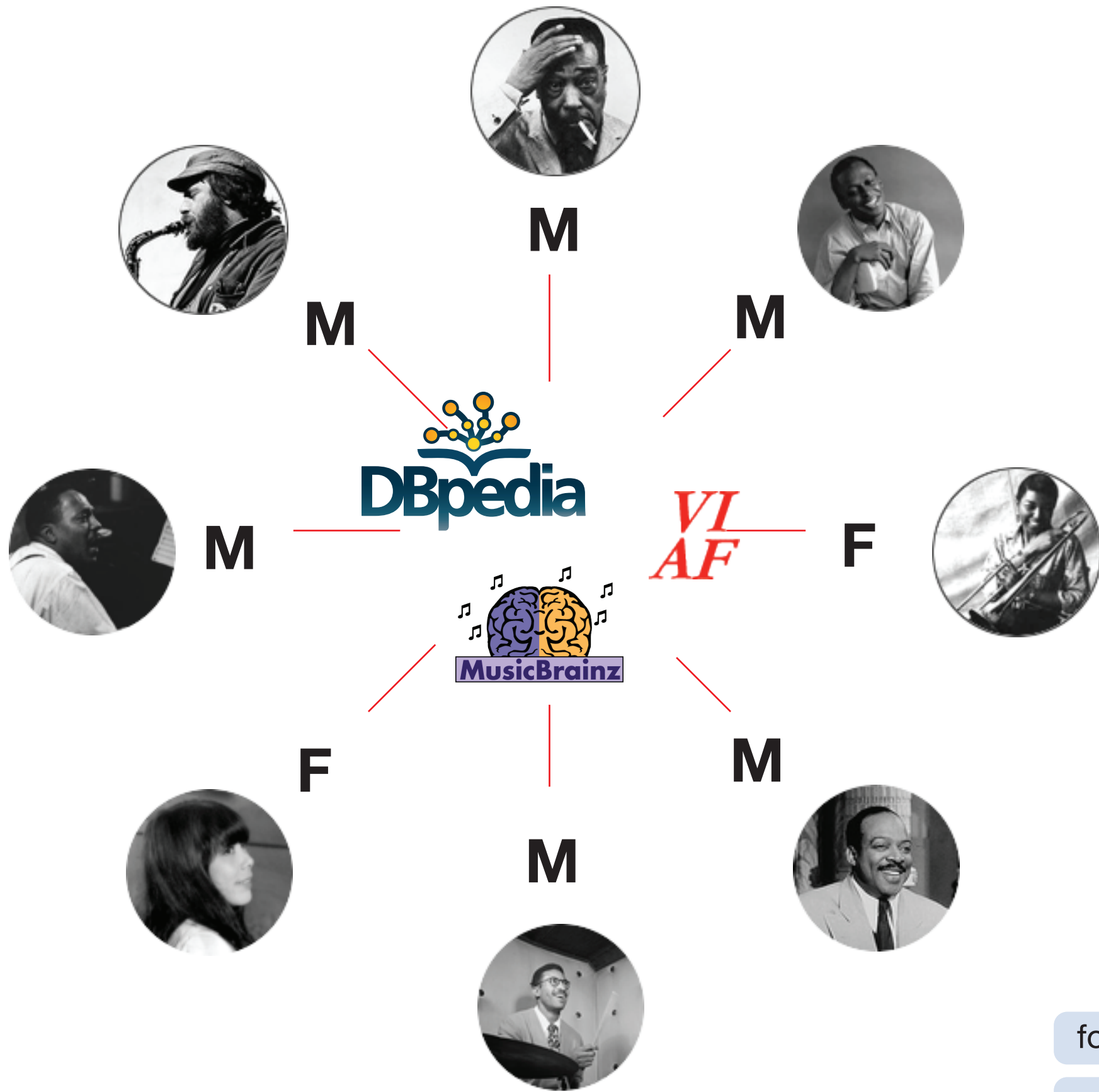
About Linked Jazz

Linked Jazz is a research project investigating the application of Linked Open Data technologies to digital cultural heritage materials. Our goal is to uncover meaningful connections between documents and data related to the personal and professional lives of musicians and to develop broadly applicable tools and methods for working with Linked Open Data. At the heart of our work are the transcripts of oral history interviews with jazz musicians.

This poster details some of the contributions of Linked Jazz student team members Karen Li-Lun Hwang, William J. Levay, and Alexandra Provo during the 2014-2015 academic year.

Merging Linked Open Data for Enrichment and Analysis

As more transcripts are processed, our dataset of interview-derived names from the world of jazz continues to grow.

Towards the end of 2014, Linked Jazz began focusing on the underrepresented careers of **women in jazz**. We started by targeting interviews with female jazz musicians found in archives. In order to open the **possibility of exploring gender relationships** with our tools, we tapped other Linked Open Data resources to enrich our dataset, using properties such as "sameAs" and Python scripts to query gender data from different resources. **DBpedia categories, MusicBrainz**, and **VIAF authority clusters** provide the semantic definition to allow automated mash-ups with our URI list of names. To date, the SPARQL endpoints and XML records of these three sources have provided gender information for 75% of our list. With more diverse resources entering the Linked Open Data cloud every year, researchers will be able to use tools like Linked Jazz to create datasets derived from archival materials and merge them with any open data from the cloud, offering a powerful environment for researchers to explore.

http://dbpedia.org/page/Toshiko_Akiyoshi

dbpedia-owl:viafId	14957360
owl:sameAs	fbase:Toshiko Akiyoshi
	http://cs.dbpedia.org/resource/Toshiko_Akijoši
	http://de.dbpedia.org/resource/D437117
	http://wikidata.dbpedia.org/resource/D437117
	http://cs.dbpedia.org/resource/Toshiko_Akijoši
	http://ja.dbpedia.org/resource/穐吉敏子
	http://ja.dbpedia.org/resource/秋吉敏子
	http://pt.dbpedia.org/resource/Toshiko_Akiyoshi
	http://wikidata.org/entity/Q437117
	http://yago-knowledge.org/resource/Toshiko_Akiyoshi
	http://zitgist.com/music/artist/78fe2146-ec71-44c8-86eb-117607e53933
	http://eo.dbpedia.org/resource/Toshiko_Akiyoshi
	http://fi.dbpedia.org/resource/Toshiko_Akiyoshi

INTERVIEW WITH TOSHIKO AKIYOSHI BY ANTHONY BROWN, 2008
(SMITHSONIAN JAZZ ORAL HISTORY PROGRAM NEA JAZZ MASTERS)

Anthony Brown (Interviewer): I think what Im trying to highlight here is that youre composing, so, several compositions, but well talk about that later in the interview, your process of composition, et cetera. But the fact that, you know, you have several originals on this session, so....

Yeah, right. I think its, I started writing because, when I came to Boston I used to, of course, I used to go every night to listen, and in those days like, even the big bands, like Dukes band or Count Basie band, would stay for two weeks, you know, at a club. You cant think of that today, but.... I was always sitting in, I decided that when I, if somebody asked me to sit in, I was going to sit in. You know, I decided that. So, I used to sit in with, Max (Roach) would let me sit in, Miles Davis would let me sit in, so on, so on. So ... Diz had a big band in those days, and Phil Woods was, I think, lead alto, and uh, Melba Liston was playing trombone. Anyway, uh, I was sitting in, playing, and I heard, "Bud Powell! Bud Powell!" When I played people said, "Aw, shes Bud Powell! Shes Bud Powell!" And I started thinking, I said, "I guess thats no good, I have to, you know, to try to find my own idiosyncrasy. " One of the (reasons) I started writing, one of the things that, I think, thats the reason I started writing, to develop my own idiosyncrasy. So writing was something I was never interested in before.

Connecting Musicians Through the Photo Archive

Tulane University has a rich collection of historical photographs of jazz musicians living and performing in New Orleans and around the world. We received two tab-delimited text files from Tulane, metadata from **The Hogan Jazz Archive Photography Collection** and the **Ralston Crawford Collection of Jazz Photography**, exported from their CONTENTdm system.

Some statistics: this dataset contains **1,787 images**, at least **681 unique individuals**, and more than **2,700 depictions**. Depiction is the FOAF term that we later used as a predicate in our triples from this dataset. One group photograph might depict several individuals, and one individual might be depicted in several photographs. People depicted in the same photograph might be said to "know" each other in some way.

Using Open Refine and Python scripts, we standardized the photograph metadata, reconciled names with VIAF, then described the photographs, and the people and relationships depicted, using RDF triples.

Title	Louis Armstrong's Hot Five
Creator	Daguerre, Chicago
Subject	St. Cyr, Johnny, 1890-1966 Ory, Edward Armstrong, Louis, 1901-1971 Dodds, Johnny, 1892-1940 Armstrong, Lil Hardin, 1902-1971 Publicity photographs

```
<http://dbpedia.org/resource/Lil_Hardin_Armstrong>
  <http://xmlns.com/foaf/0.1/depiction>
    <http://cdm16313.contentdm.oclc.org/cdm/ref/collection/p
16313coll133/id/45>

<http://dbpedia.org/resource/Lil_Hardin_Armstrong>
  <http://xmlns.com/foaf/0.1/knows>
    <http://dbpedia.org/resource/Johnny_St._Cyr>
    <http://dbpedia.org/resource/Kid_Ory>
    <http://dbpedia.org/resource/Louis_Armstrong>
    <http://dbpedia.org/resource/Johnny_Dodds>

<http://cdm16313.contentdm.oclc.org:80/cdm/ref/collection/
p16313coll133/id/45>
  <http://purl.org/dc/terms/spatial>
    <http://sws.geonames.org/4887539/>
  <http://purl.org/dc/terms/created>
    "1926"^^<http://www.w3.org/2001/XMLSchema#gYear>
```

RDF EXAMPLES DERIVED FROM THE TULANE DATASET AND A SAMPLE NETWORK GRAPH SHOWING SOME OF THESE RELATIONSHIPS

foaf:depiction
foaf:knows

Refine
python
VIAF

Performance History Data & the Linked Jazz Ontology

The professional relationships between working jazz musicians is a central focus of the Linked Jazz project. Data about concert performances and recording dates gives us rich information about not just collaborations between musicians, but also about time and place, musical works, songs, and songwriters, and record labels and releases — **all potential access points for users and researchers**.

The Carnegie Hall Archives has a rich collection of concert programs, and archivist Rob Hudson has shared with us this performance history data expressed as RDF triples. We are also working with the author of BRIAN, a relational database application for compiling recording session-based discographies, to translate **user-generated jazz discographies** into linked data.

As we begin ingesting these new datasets to produce mashups, the need for a more robust, domain-specific ontology has arisen. During the development of our **52nd Street crowdsourcing tool** we created five new predicates describing relationships between musicians. Now, using the data model produced for the Carnegie Hall and BRIAN collaborations as a starting point, Linked Jazz has drafted an expanded inventory of RDF classes and predicates. We've begun to model new entities like **music groups** and **images** using RDFS (RDF Schema) and OWL.

foaf:Agent
foaf:Group   mo:MusicArtist
mo:MusicGroup
foaf:member
mo:member
rel:influencedBy
mo:collaborated_with
mo:member_of
lj:playedTogether
lj:touredWith

DATASET MAPPING SPREADSHEET, MUSIC GROUP DIAGRAM (ABOVE), AND IMAGE DESIGN (LEFT)

DESIGN OF THE CLASS IMAGE

```
owl:Thing rdf:type rdfs:Class
  foaf:Image rdf:type rdfs:Class
    owl:equivalentClass
    schema:ImageObject rdf:type rdfs:Class

Properties:
  foaf:depiction
    Domain:    owl:Thing
    Range:     foaf:Image
  foaf:depicts
    Domain:    foaf:Image
    Range:     owl:Thing
  foaf:thumbnail
    owl:equivalentProperty
  schema:thumbnail
    Domain:    foaf:Image
    Range:     schema:ImageObject
    Domain:    foaf:Image
    Range:     schema:ImageObject
  schema:image
    Domain:    schema:ImageObject
    Range:     schema:Thing
```
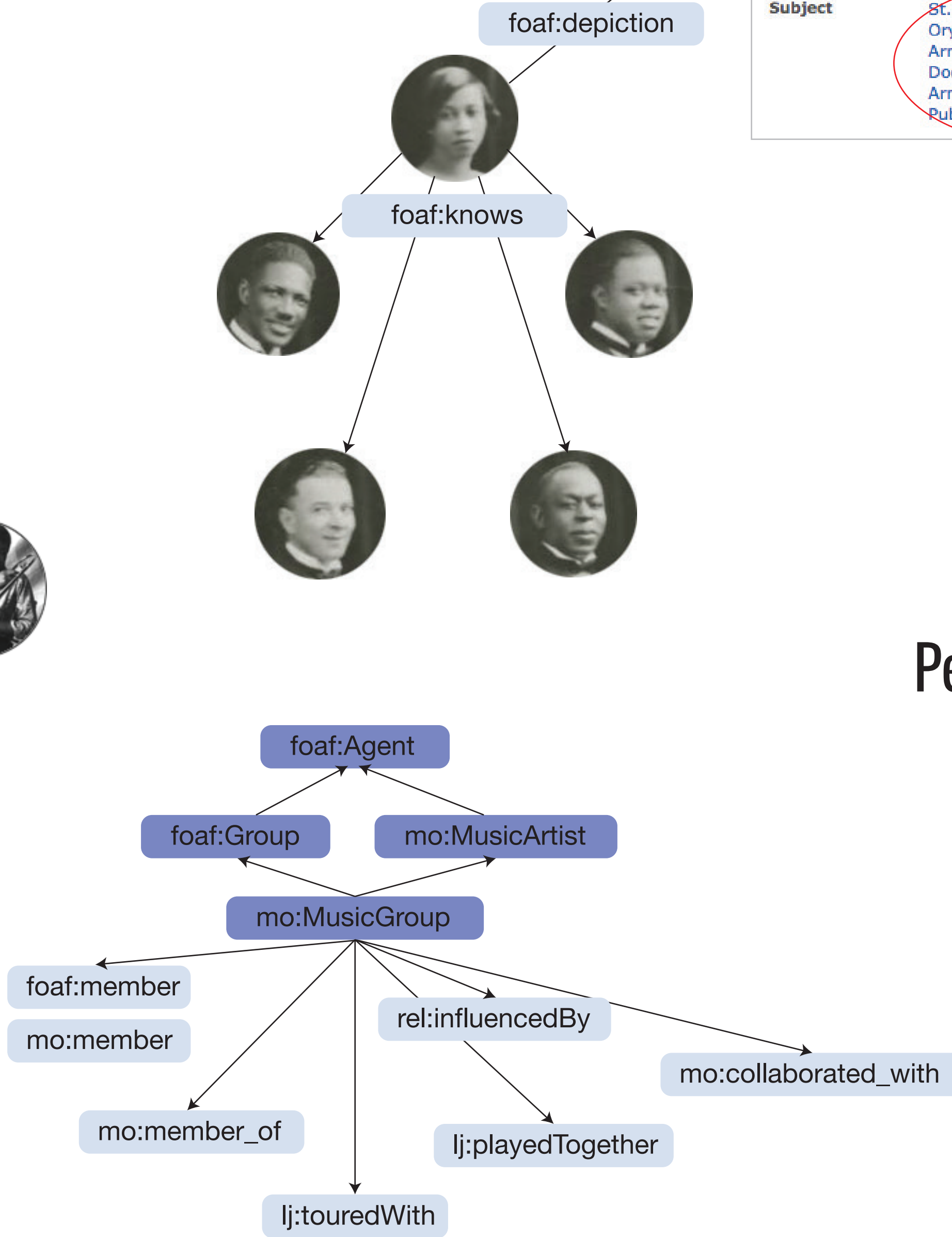
Person or Group	Linked Jazz	Carnegie Hall	
ID			id
Person or Group	foaf:Person	rdf:type	
Name: Last First			
Name: Last			last name
Name: First			first name
Name: Full	foaf:name	foaf:name	full name
Name: Alt			
Name: Suffix			qualifier
URI		owl:sameAs	dbpedia
		owl:sameAs	lcnaf
		owl:sameAs	mbz
Gender			
Date of Birth		schema:birthDate	birth date (YYYY-MM-DD)
Date of Death		schema:deathDate	death date (YYYY-MM-DD)
Place Active			
Birthplace			birth city
			birth state
			birth country
Death place		dbpedia-owl:birthPlace	death city
			death state
			death country
			geoDeathPlace
Ethnicity			
Instrument		http://d-nb.info/standards/elementset/gnd#playedInstrument	instrument
Band or group			[group]
Occupation		http://d-nb.info/standards/elementset/gnd#professionOrOccupation	this is an array of occupations
Notes			notes
Thumbnail	foaf:depiction		
Relationship	[list of relationship predicates]		

With thanks to Cristina Pattuelli, Matt Miller, the Linked Jazz team, and our collaborators. Special thanks to the musicians of jazz.