

# **Data for a Post-Silo Age Part 2**

## **Linked Data Integration**

**Pratt School of Information ASIS&T Speakeasy  
Hannah Sistrunk and Molly Reese-Lerner  
March 23, 2016**

# Linked Open Data



**Imagining how can we move  
beyond our own network  
and integrate with outside data...**

# Linked Jazz Meets Carnegie Hall

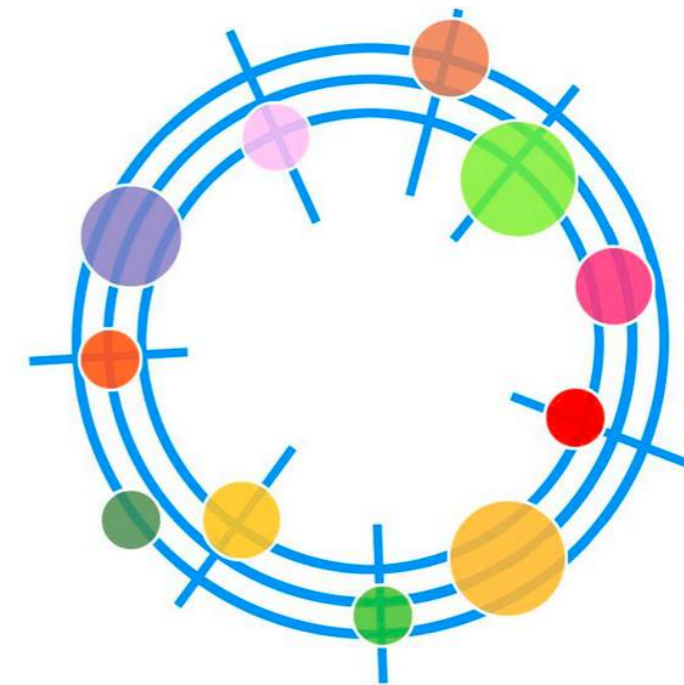
## Pilot study goals:

- Interlinking: identifying Carnegie Hall performers in Linked Jazz network
- Visualization of shared relationships among the matches
- Explore possibilities and identify problems in data integration process

**CARNEGIE HALL**



**LINKED JAZZ** 



## Performance History Database

- **90,000 performers**
- **50,000 events**
- **16,000 composers**
- **85,000 musical works**

## Jazz Transcript Data

- **2006 people**
- **~9000 "knows of" relationships**

# Data Subset - Carnegie Hall Jazz Events

- All 'jazz' events at Carnegie Hall from 1912 to 1955
- Date, time, title, venue (specific hall), performers, groups
- Sources include programs, concert calendars, newspapers



# Carnegie Hall Event Triples

```

<http://data.carnegiehall.org/events/CH_19460104_2030> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Al_Lucas_(musician)> .
<http://data.carnegiehall.org/events/CH_19470316_0000> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Lionel_Hampton> .
<http://data.carnegiehall.org/events/CH_19440402_2030> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Mezz_Mezzrow> .
<http://data.carnegiehall.org/events/CH_19431211_2045> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Fred_Guy> .
<http://data.carnegiehall.org/events/CH_19490220_2030> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Symphony_Sid> .
<http://data.carnegiehall.org/events/CH_19460603_2030> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Illinois_Jacquet> .
<http://data.carnegiehall.org/events/CH_19441225_1730> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Sammy_Price> .
<http://data.carnegiehall.org/events/CH_19490918_0000> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Hank_Jones> .
<http://data.carnegiehall.org/events/CH_19550316_2030> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Duke_Ellington> .
<http://data.carnegiehall.org/events/CH_19471129_2030> <http://purl.org/NET/c4dm/event.owl#place> <http://data.carnegiehall.org/venues/Main_Hall> .
<http://data.carnegiehall.org/events/CH_19481225_2030> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Dizzy_Gillespie> .
<http://data.carnegiehall.org/events/CH_19510221_2030> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Gene_Ramey> .
<http://data.carnegiehall.org/events/CH_19120502_2015> <http://purl.org/ontology/mo/performer> <http://data.carnegiehall.org/names/Charlie_Mills> .
<http://data.carnegiehall.org/events/CH_19470316_0000> <http://purl.org/ontology/mo/performer> http://dbpedia.org/resource/Ella_Fitzgerald .
<http://data.carnegiehall.org/events/CH_19211006_2030> <http://purl.org/ontology/mo/performer> <http://data.carnegiehall.org/names/Jimmy_Abato> .
<http://data.carnegiehall.org/events/CH_19120502_2015> <http://purl.org/ontology/mo/performer> <http://data.carnegiehall.org/names/Elizabeth_Payne> .
<http://data.carnegiehall.org/events/CH_19550213_0000> <http://purl.org/ontology/mo/performer> <http://dbpedia.org/resource/Count_Basie> .
<http://data.carnegiehall.org/events/CH_19480416_2000> <http://www.w3.org/2000/01/rdf-schema#label> "Benefit: Sydenham Hospital"@en .
<http://data.carnegiehall.org/events/CH_19550403_0000> <http://purl.org/ontology/mo/performer> http://dbpedia.org/resource/Pearl_Bailey .
<http://data.carnegiehall.org/events/CH_19510928_1945> <http://www.w3.org/2000/01/rdf-schema#label> "Jazz Concert featuring Duke Ellington and Others"@en .
    
```

**/CH\_19470316\_2030**

**Subject**

**/performer**

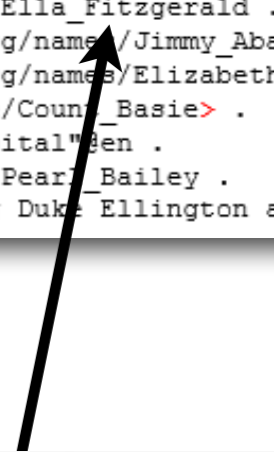
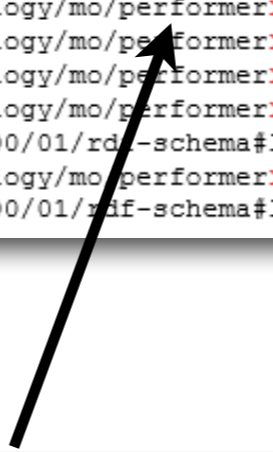
**Predicate**

**/Ella\_Fitzgerald**

**Object**

-

-



# Linked Jazz People Triples

```

<http://dbpedia.org/resource/John_Coltrane> <http://xmlns.com/foaf/0.1/name> "John Coltrane"@en .
<http://dbpedia.org/resource/Dizzy_Gillespie> <http://xmlns.com/foaf/0.1/name> "Dizzy Gillespie"@en .
<http://dbpedia.org/resource/Quincy_Jones> <http://xmlns.com/foaf/0.1/name> "Quincy Jones"@en .
<http://dbpedia.org/resource/Wynton_Marsalis> <http://xmlns.com/foaf/0.1/name> "Wynton Marsalis"@en .
<http://dbpedia.org/resource/Miles_Davis> <http://xmlns.com/foaf/0.1/name> "Miles Davis"@en .
<http://dbpedia.org/resource/Billie_Holiday> <http://xmlns.com/foaf/0.1/name> "Billie Holiday"@en .
<http://dbpedia.org/resource/Rodney_Kendrick> <http://xmlns.com/foaf/0.1/name> "Rodney Kendrick"@en .
<http://dbpedia.org/resource/Thad_Jones> <http://xmlns.com/foaf/0.1/name> "Thad Jones"@en .
<http://dbpedia.org/resource/Johnny_Griffin> <http://xmlns.com/foaf/0.1/name> "Johnny Griffin"@en .
<http://dbpedia.org/resource/Lionel_Hampton> <http://xmlns.com/foaf/0.1/name> "Lionel Hampton"@en .
<http://dbpedia.org/resource/Gene_Ammons> <http://xmlns.com/foaf/0.1/name> "Gene Ammons"@en .
<http://dbpedia.org/resource/Illinois_Jacquet> <http://xmlns.com/foaf/0.1/name> "Illinois Jacquet"@en .
<http://dbpedia.org/resource/Joe_Williams_(jazz_singer)> <http://xmlns.com/foaf/0.1/name> "Joe Williams"@en .
<http://dbpedia.org/resource/James_Moody_(saxophonist)> <http://xmlns.com/foaf/0.1/name> "James Moody"@en .
<http://dbpedia.org/resource/Ella_Fitzgerald> <http://xmlns.com/foaf/0.1/name> "Ella Fitzgerald"@en .
<http://dbpedia.org/resource/Charles_Davis_(flute_player)> <http://xmlns.com/foaf/0.1/name> "Charles Davis"@en .
<http://dbpedia.org/resource/Sonny_Rollins> <http://xmlns.com/foaf/0.1/name> "Sonny Rollins"@en .
<http://dbpedia.org/resource/Sun_Ra> <http://xmlns.com/foaf/0.1/name> "Sun Ra"@en .
    
```

**/Ella\_Fitzgerald**

**/name**

**"Ella Fitzgerald"**

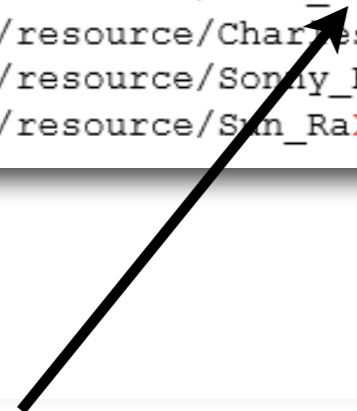
**Subject**

-

**Predicate**

-

**Object**



# Step 1: Finding People Matches

Find matching DBpedia URIs between datasets:

LINKED JAZZ 

`<http://dbpedia.org/resource/Ella_Fitzgerald>`

`<http://xmlns.com/foaf/0.1/name>`

`"Ella Fitzgerald"@en.`

CARNEGIE HALL

`<http://data.carnegiehall.org/events/CH_19470316_0000>`


`<http://purl.org/ontology/mo/performer>`

`<http://dbpedia.org/resource/Ella_Fitzgerald>`


**=268 matches!**



## Not all performers in Carnegie Hall event triples have URIs that can be matched to Linked Jazz

CARNEGIE HALL `<http://data/carnegiehall.org/names/Ralph_Cooper>`  
=  `<http://dbpedia.org/resource/Ralph_Cooper_(Apollo)>`

---

CARNEGIE HALL  
`<http://data.carnegiehall.org/names/Fats_Waller_His_Orchestra>`  
= `<http://dbpedia.org/resource/Fats_Waller>`  
 LINKED JAZZ

## Use Carnegie Hall "address book" data describing ALL performers (74,000+ triples)

-dates of birth, places of birth, instruments, professions, etc.

This data includes name in "string" form (not URI):

CARNEGIE HALL

```
<http://data.carnegiehall.org/names/Ralph_Cooper>
```

```
<http://xmlns.com/foaf/0.1/name> "Ralph Cooper" .
```

```
<http://dbpedia.org/resource/Ralph_Cooper_(Apollo)>
```

```
<http://xmlns.com/foaf/0.1/name> "Ralph Cooper"@en .
```

LINKED JAZZ 

### Python Human Name Parser...

```
{  
  "full name": "Willie &quot;the Lion&quot; Smith",  
  "first name": "Willie",  
  "last name": "Smith"  
}
```

**=264 matches!**

# Data Clean-up

**Removing duplicates, total matches  
= 373**

**New dataset of performers with their  
associated URIs from Linked Jazz**

# Step 2: From matches, pull "knows of" relationships Linked Jazz data

```
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Max_Roach> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://linkedjazz.org/resource/Harris_Levy> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://id.loc.gov/authorities/names/n83056867> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Kenny_Clarke> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://linkedjazz.org/ontology/playedTogether> <http://dbpedia.org/resource/Kenny_Clarke> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Bill_Coleman> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Billy_Taylor> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Al_Hall_(musi>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Kai_Winding> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Percy_Heath> .
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Woody_Hermap>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Mickey_Rokej>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Kenny_Dorhan>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Kenny_Dorhan>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://musicbrainz.org/artist/10a7aaa5-5>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://linkedjazz.org/resource/Father_Ar>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://musicbrainz.org/artist/1387e590-7>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/hasMet> <http://musicbrainz.org/artist/1387e590-7c>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Lena_Horne>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Bary_Ulanov>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://linkedjazz.org/resource/Father_Cj>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://linkedjazz.org/resource/Boo_Frazi>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://musicbrainz.org/artist/d8d94cfc-5>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Melba_Liston>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://linkedjazz.org/resource/Sy_Baron>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/hasMet> <http://linkedjazz.org/resource/Sy_Baron>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://musicbrainz.org/artist/bdd5ffb6-5>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://id.loc.gov/authorities/names/nr00>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Rogers_White>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Fred_Robbins>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Babs_Gonzalez>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://linkedjazz.org/resource/Peggy_Mel>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://dbpedia.org/resource/Katherine_Dr>
<http://dbpedia.org/resource/Mary_Lou_Williams> <http://purl.org/vocab/relationship/knowsOf> <http://id.loc.gov/authorities/names/n926>
```

CSV

Mary Lou Williams	Fats Navarro
Mary Lou Williams	Blanche Calloway
Mary Lou Williams	David Stone Martin
Mary Lou Williams	Jo Jones
Mary Lou Williams	Stumpy Brady
Mary Lou Williams	Alvin Ailey
Mary Lou Williams	Lena Horne
Mary Lou Williams	Kenny Durham
Mary Lou Williams	Johnny Hodges
Mary Lou Williams	Van Cliburn
Mary Lou Williams	Melba Liston
Mary Lou Williams	Johnny Garrett
Mary Lou Williams	Al Hall
Mary Lou Williams	Sarah Vaughan
Mary Lou Williams	Hot Lips Page
Mary Lou Williams	Ted Heath
Mary Lou Williams	George Wein
Mary Lou Williams	J. C. Heard
Mary Lou Williams	Milton Orent
Mary Lou Williams	Peggy Mellon Hitchcock
Mary Lou Williams	Willis Scruggs
Mary Lou Williams	Edmond Hall
Mary Lou Williams	Miles Davis
Mary Lou Williams	Johnny Dankworth
Mary Lou Williams	Howard Burley
Mary Lou Williams	Joe Smith
Mary Lou Williams	Joe Turner
Mary Lou Williams	Earl Hines

/Mary\_Lou\_Williams

/knowsOf

/Lena\_Horne

Subject

-

Predicate

-

Object

From RDF triples to CSV using Python

# Step 3: Create relationships for people who played in the same CH event together

## Event-performer triple:

```
<http://data.carnegiehall.org/events/CH_19490211_2330>  
<http://purl.org/ontology/mo/performer>  
<http://dbpedia.org/resource/Ella_Fitzgerald> .
```

## Event-performer dictionary:

```
"http://data.carnegiehall.org/events/CH_19490211_2330": [  
  "http://dbpedia.org/resource/Ella_Fitzgerald",  
  "http://dbpedia.org/resource/Coleman_Hawkins",  
  "http://dbpedia.org/resource/Ray_Brown_(musician)",  
  "http://dbpedia.org/resource/Shelly_Manne",  
  "http://dbpedia.org/resource/Flip_Phillips",  
  "http://dbpedia.org/resource/Hank_Jones",  
  "http://dbpedia.org/resource/Fats_Navarro",  
  "http://dbpedia.org/resource/Charlie_Parker"  
],
```

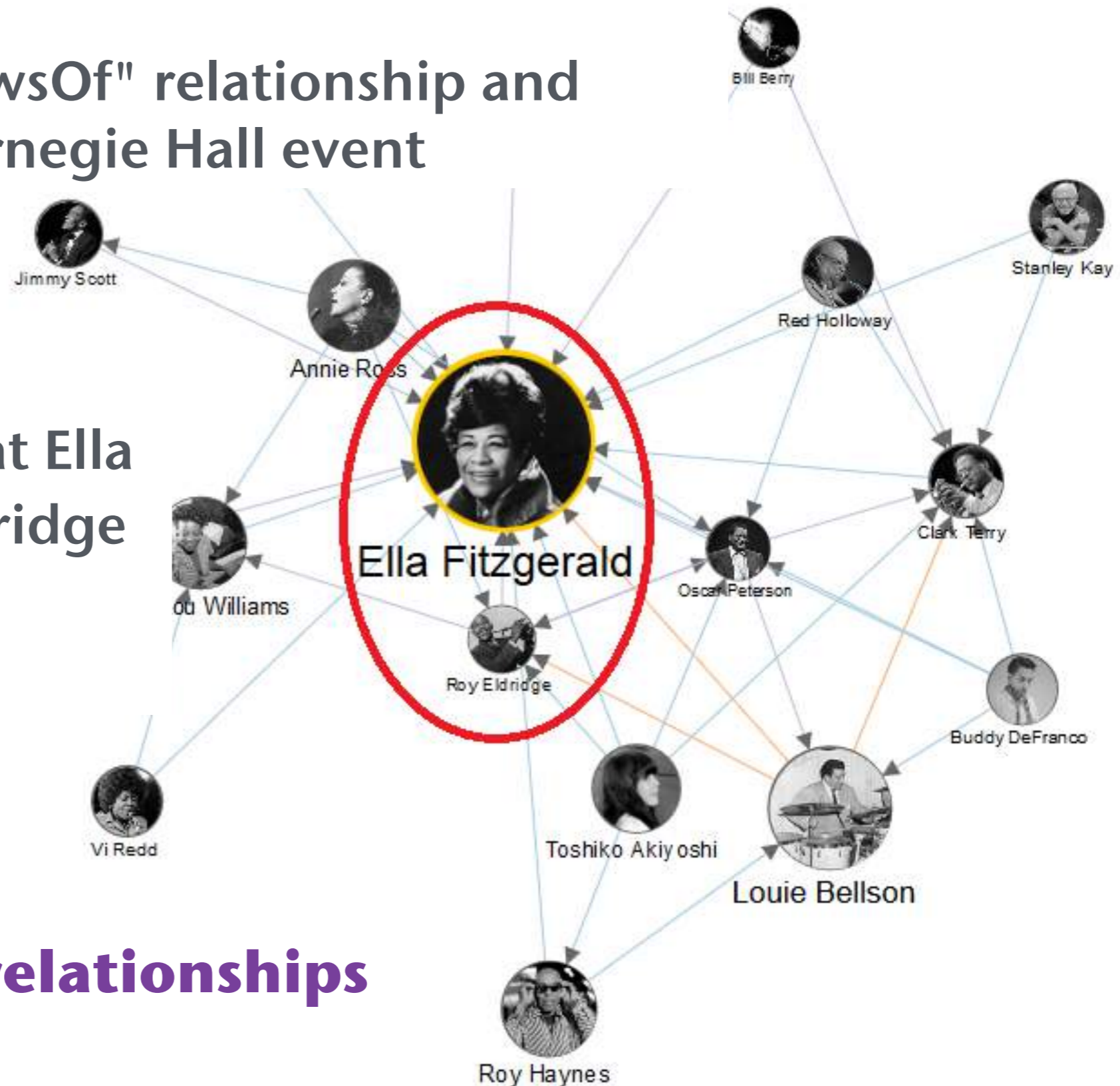
## Performer relationships:

Ella Fitzgerald	Oscar Peterson
Ella Fitzgerald	Ray Brown
Ella Fitzgerald	Charlie Shavers
Ella Fitzgerald	Herb Ellis
Ella Fitzgerald	J. C. Heard
Ella Fitzgerald	Benny Carter
Ella Fitzgerald	Roy Eldridge
Ella Fitzgerald	Flip Phillips
Ella Fitzgerald	Lester Young
Ella Fitzgerald	Gene Krupa
Ella Fitzgerald	Ben Webster
Ella Fitzgerald	Shadow Wilson
Ella Fitzgerald	Joe Newman
Ella Fitzgerald	Illinois Jacquet
Ella Fitzgerald	Buddy Rich
Ella Fitzgerald	Hank Jones
Ella Fitzgerald	Barney Kessel
Ella Fitzgerald	Milt Jackson
Ella Fitzgerald	Dizzy Gillespie
Ella Fitzgerald	Charlie Parker
Ella Fitzgerald	James Moody
Ella Fitzgerald	Al McKibbin
Ella Fitzgerald	Chano Pozo
Ella Fitzgerald	Leonard Feather
Ella Fitzgerald	Woody Herman
Ella Fitzgerald	Frankie Laine
Ella Fitzgerald	Louis Jordan
Ella Fitzgerald	Lennie Tristano
Ella Fitzgerald	Bud Powell
Ella Fitzgerald	Lee Konitz
Ella Fitzgerald	Ralph Burns
Ella Fitzgerald	Kai Winding

# Step 4: Identify relationships that exist in both datasets

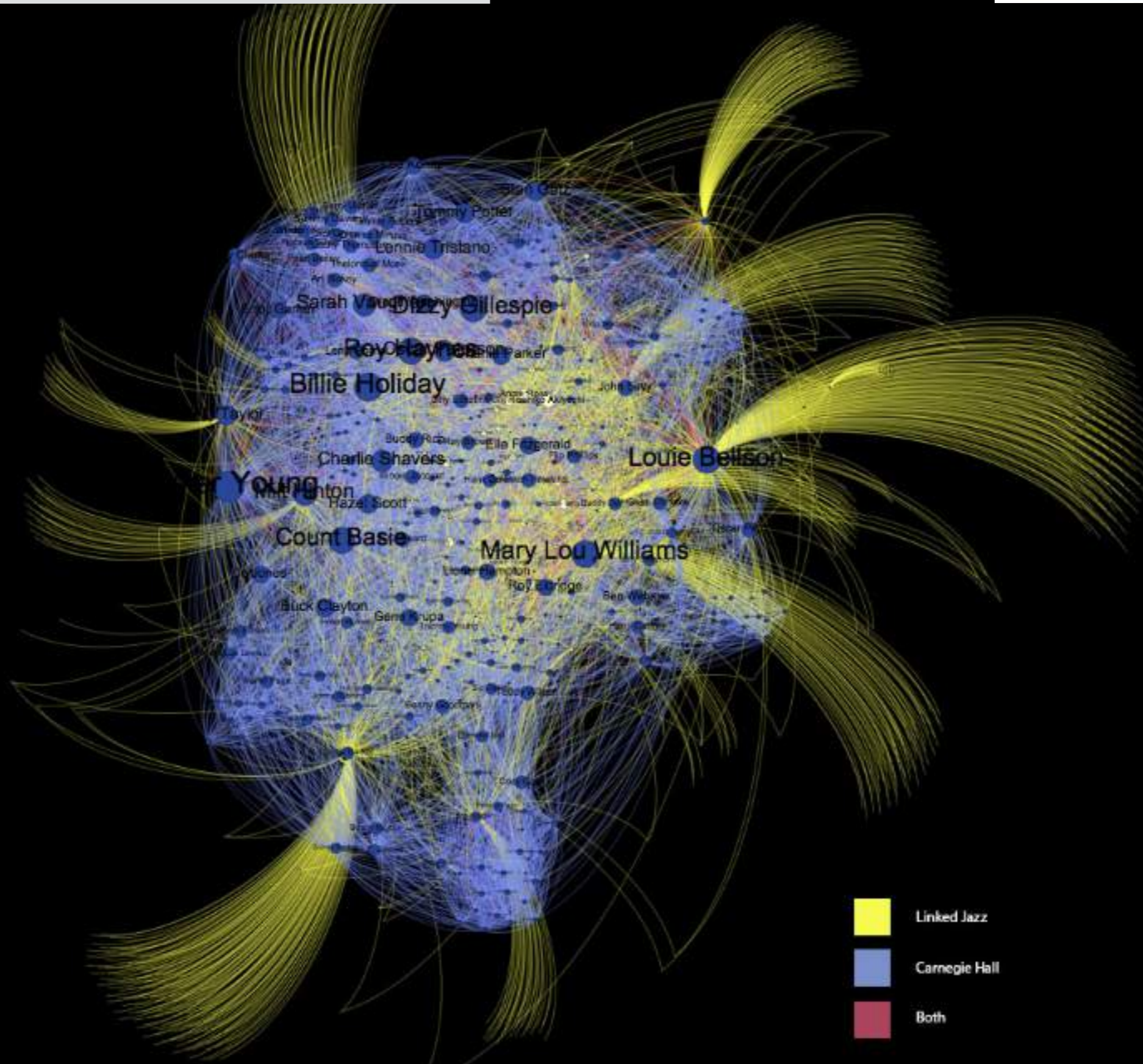
- Both Linked Jazz "knowsOf" relationship and played at the same Carnegie Hall event

- The CH data shows that Ella Fitzgerald and Roy Eldridge performed at a 1953 Carnegie Hall event



**Total = 293 shared relationships**

# Gephi Visualization



[http://pfch.nyc/linked\\_jazz\\_meets\\_carnegie\\_hall/CH-LJ\\_network/](http://pfch.nyc/linked_jazz_meets_carnegie_hall/CH-LJ_network/)

# Next Steps

```
<http://data.carnegiehall.org/events/CH_19490211_2330>  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
<http://purl.org/NET/c4dm/event.owl#Event>
```

→ **Type: Event**

```
<http://data.carnegiehall.org/events/CH_19490211_2330>  
<http://www.w3.org/2000/01/rdf-schema#label>  
"Jazz at the Philharmonic"@en
```

→ **Label: "Jazz at the Philharmonic"**

```
<http://data.carnegiehall.org/events/CH_19490211_2330>  
<http://purl.org/NET/c4dm/event.owl#place>  
<http://data.carnegiehall.org/venues/Main_Hall>
```

→ **Venue: Main Hall**

```
<http://data.carnegiehall.org/events/CH_19490211_2330>  
<http://purl.org/dc/terms/date> "1949-02-11T23:30:00-05:00"^^  
<http://www.w3.org/2001/XMLSchema#dateTime>
```

→ **Date: Feb. 11, 1949**

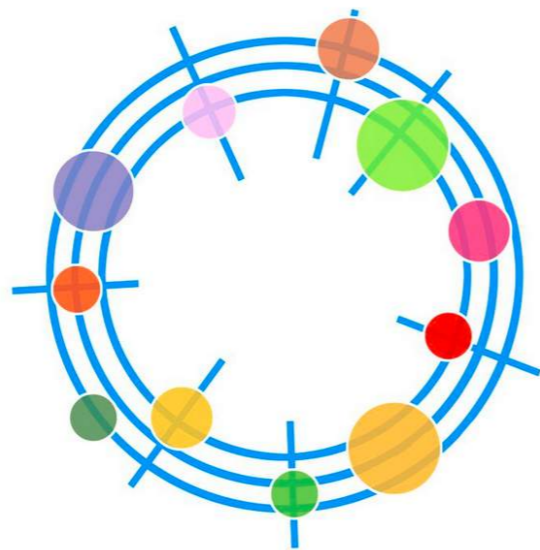
```
<http://data.carnegiehall.org/events/CH_19480413_2030>  
<http://purl.org/ontology/mo/performer>  
<http://dbpedia.org/resource/Ella_Fitzgerald>
```

→ **Performer/s**



# Conclusion

- Learn by doing! Python, SPARQL, data wrangling, trial and error
- Best practices not well established
  - develop methodologies
  - contribute to community by sharing research
- Get involved with the LOD Research Group at Pratt



website: [linkedjazz.org](http://linkedjazz.org)

email: [linkedjazz@linkedjazz.org](mailto:linkedjazz@linkedjazz.org)

twitter: [@linkedjazz](https://twitter.com/linkedjazz)

# Acknowledgements

**Thank you!**

- Project directors Dr. Cristina Pattuelli and Matt Miller
- Rob Hudson - Associate Archivist, Carnegie Hall
- William Levay and Molly Seegers