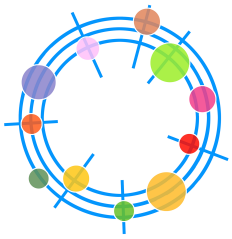


Linked Jazz

Revealing the relationships of the jazz community

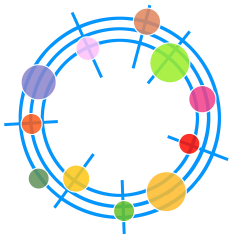
Matt Miller
@thisismiller



Project Overview

- Investigating the application of Linked Open Data to enhance the discovery and visibility of digital cultural heritage materials.
- Build new methods of connecting cultural data.
- Uncover meaningful connections between documents and data related to the personal and professional lives of musicians who often practice in rich and diverse social networks.

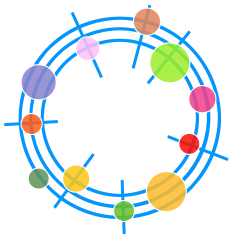
Professor Cristina Pattuelli at the Pratt Institute School of Library Information Science is the director of the project which began in 2011.



Linked Data Now!

Why?

- Bootstrap your project with existing data.
- Highlights knowledge you have created and knowledge that is missing.
- Facilitates sharing, but also growing your own project.

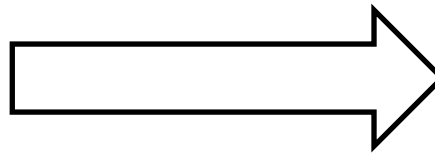


Bootstrapping – Identifying



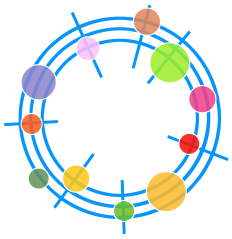
Research Question

How can we discover and analyze the rich and diverse network of relationships between jazz musicians?



Primary Sources

Oral history interview transcripts of jazz musicians.



Bootstrapping – Identifying



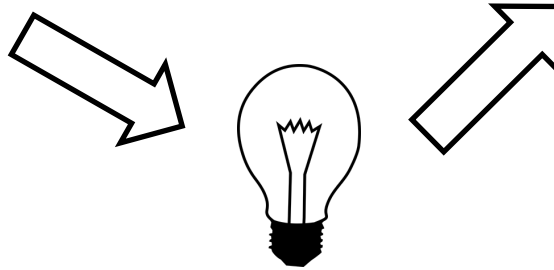
Research Question

How can we discover and analyze the rich and diverse network of relationships between jazz musicians?

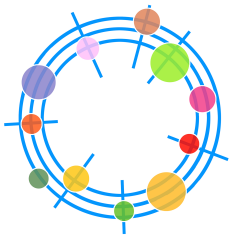


Primary Sources

Oral history interview transcripts of jazz musicians.



We need to know the names (and variants) of jazz musicians in a structured controlled vocabulary.



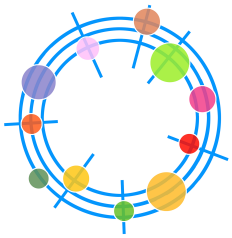
Bootstrapping – Identifying



Charlie Parker

```
<skos:prefLabel>Parker, Charlie, 1920-1955</skos:prefLabel>  
<skos:altLabel>Yardbird.</skos:altLabel>  
<skos:altLabel>Bird.</skos:altLabel>  
<skos:altLabel>Parker, Charles Christopher.</skos:altLabel>  
<skos:altLabel>Parker, Charles.</skos:altLabel>  
<skos:altLabel>Charlie Chan.</skos:altLabel>
```

Many different LOD datasets contain this information. We need to access, query and link it for only jazz related individuals.

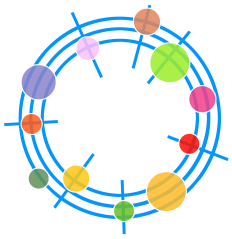


Bootstrapping – Querying



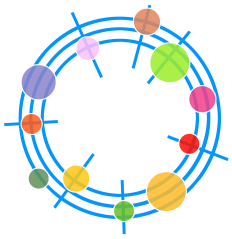
Background information	
Birth name	Charles Parker, Jr.
Also known as	Bird, Yardbird, Zoizeau (in France) ^[1]
Born	August 29, 1920 Kansas City, Kansas, United States
Died	March 12, 1955 (aged 34) New York City, New York, United States
Genres	Jazz, bebop
Occupations	Saxophonist, composer
Instruments	Alto saxophone, tenor saxophone
Years active	1937–1955
Labels	Savoy, Dial, Verve
Associated acts	Miles Davis, Max Roach
Website	www.cmgww.com/music/parker/

Categories: [1920 births](#) | [1955 deaths](#) | [20th-century American musicians](#) | [20th-century composers](#) | [American buskers](#) | [American jazz composers](#) | [American jazz saxophonists](#) | [Bebop saxophonists](#)



Bootstrapping – Querying

- Processing the DBpedia dataset resulted in around 9,000 URIs.
 - DBpedia is fluid! After each release (currently 3.9) we reprocess the files resulting in the addition of 500-700 URIs.
- We now have a name directory, but we want additional forms of personal names. To accomplish this we try mapping to Library of Congress.
- Matching DBpedia and LC URIs is not automatic.

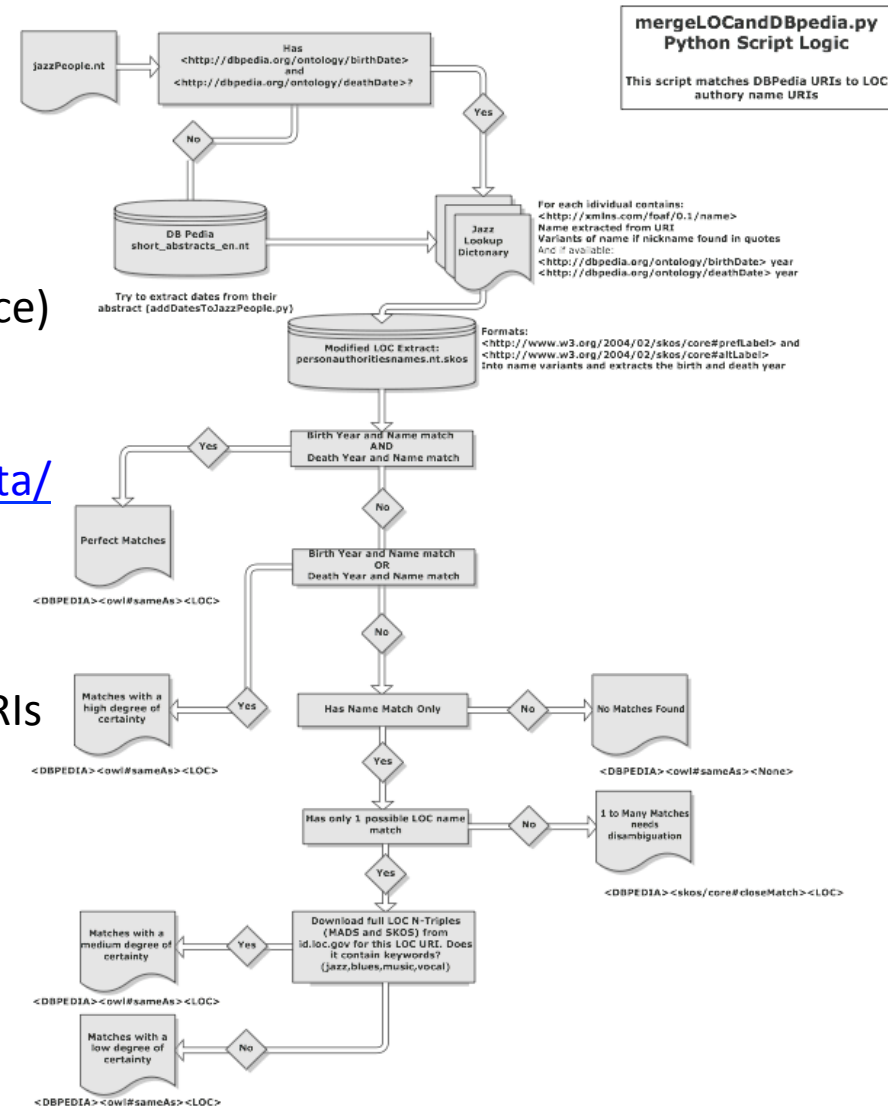


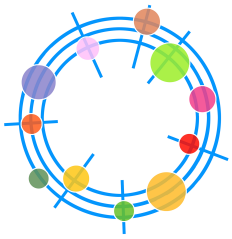
Bootstrapping – Mapping

- We matched identities based on:
 - Name
 - Life Dates
 - White listed words found in sources (<http://www.loc.gov/mads/rdf/v1#Source>)
- Reconciling authorities is difficult!
 - Use others work: <http://viaf.org/viaf/data/>
- But don't discount your own processes.
 - Using our relatively simple process we were able to match about 1500 more URIs than VIAF.org.
 - This is due to a smaller domain (jazz).

Our name directory creation and authority matching is documented:

<https://github.com/thisismattmiller/linked-jazz-name-directory>





Bootstrapping – Curating

LINKED JAZZ 

Perfect (982) **High (1,024)** Medium (2,597) Low (1,463) Many (1,116) None (2,117) | Verified (17) Deleted(1) | 

Mikhail Alperin	1956 -			VI AF			
Trigger Alpert	1916 -			VI AF			
John Altman	1949 -			VI AF			
Ray Anderson	1952 -			VI AF			
Reid Anderson	1970 -			VI AF			
Eddie Anthony	1890 - 1934			VI AF			
Ray Anthony	1922 -			VI AF			
Marc Antoine	1963 -			VI AF			
Peter Appleyard	1928 -			VI AF			
Louis Armstrong	1900 - 1971			VI AF			
Leo Arnaud	1904 - --04			VI AF			
Horace Arnold	1937 -			VI			

Louis Armstrong



Louis Armstrong (August 4, 1901 – July 6, 1971), nicknamed Satchmo or Pops, was an American jazz trumpeter and singer from New Orleans, Louisiana. Coming to prominence in the 1920s as an "inventive" trumpet and cornet player, Armstrong was a foundational influence in jazz, shifting the

focus of the music from collective improvisation to solo performance.

http://dbpedia.org/resource/New_Orleans

<http://dbpedia.org/resource/Louisiana>

 LIBRARY OF CONGRESS

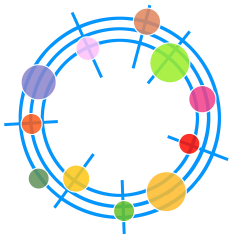
<http://id.loc.gov/authorities/names/n50001506>

altLabel | Satchmo, 1901-1971

altLabel | Armstrong, Satchmo, 1901-1971

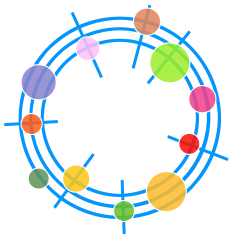
prefLabel | Armstrong, Louis, 1901-1971

http://linkedjazz.org/public_demo_mapping/



Bootstrapping – Review

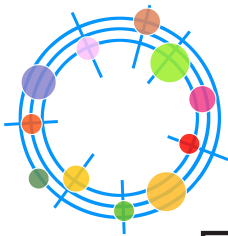
- Start small, think big.
 - Specific subject domain.
 - Large infrastructure not required (triple stores, etc.)
 - Can get started with extract files and python scripting.
- Reuse as much as possible, but try new processes leveraging domain specificity.
- Always be curating, use tools to facilitate process but a human hand is often required.



Applying the Data

- Use the name directory to locate individuals in the interview transcript.
- This project phase involves 50 transcripts.
- Because the names are tied to URIs we can infer a relationship triple between two individuals.

<foaf:Person> <rel:knowsOf> <foaf:Person>



Applying the Data

326 MR: That's great. Tell me about — I've been waiting to get to the fifteen years you spent at
327 Ryan's?

328 EL: Oh yes.

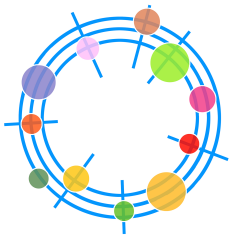
329 MR: Wow. That's a long time for a...

330 EL: Well Roy Eldridge was my — well let's see — he was my conscience.

331 MR: Yeah?


332 EL: Well he was like my — I don't know what to call him. I had a great relationship with him
333 and Coleman, but both of the relationships were very different.

362 EL: I've never played with anyone that loved to play as much as him. Never. And my greatest
363 story, every time I tell somebody this, they always, they love it, but I'm going to tell this
364 so this will be on film forever. I will never forget, we were playing in a place and there
365 was no one in the place, just like this room we're in now, with the band. We were up
366 there playing. And I was just like that [scats]. And he turned around and he leaned over
367 the drumset at me and he said "what are you doing?" And I said "well Roy" I says,
368 "there's nobody in here." He looked me right in — I mean he got closer — he said "I'm
369 here." That was the scariest thing, I mean and the way he said it, you know what I mean?
370 But it made a difference in me. He said "I'm here." Let's play. Because that's what he
371 did. I mean I've heard him play some of the greatest music I ever heard, in a room just
372 like this with nobody in it. He loved that horn. It was just like — that's why at his
373 funeral, when Dizzy said, He said "y'all gotta find something else to do now," he said
374 "because this is the only person that was ever named Jazz." He said "he's is the only one
375 who was ever named Jazz.



Transcript Analyzer


LINKED JAZZ 


 Load Transcript

 Names

 Partials



















 Others

 Meta-Structure

 Rules (32)

 Reprocess Transcript

 Publish

-   Helen Grayco
-  Helen Humes
-  J.J. Johnson
-  Jack Cole
-  Jack Lemmon
-  Jackie Mills
-  Jo Jones
-  Joe Williams
-  Kay Starr
-  Lena Horne
-  Les Hite
-  Louis Armstrong
-  Marilyn Monroe
-  Max Roach
-  Paul Humphrey
-  Red Holloway
-  Sarah Vaughan

he says "much as he plays, he can only be in one place at one time, and he can't work all the gigs all he wants so you're safe. " Which makes sense.

[Q]MR: Right.

[A]GW: Art Tatum was really a sweetheart. He was a nice man, except he didn't want to talk music, he wanted to talk baseball. You know he could tell you the average of Ty Cobb or anybody years before, you know? And how much they batted and what they did and all that. He was a good pinochle player too.

[Q]MR: Interesting. You have accompanied along the way some great singers, too.

[A]GW: Yeah, I've been with Lena Horne, Dinah Washington, Kay Starr, I even was with Dinah Shore one night. And let's see who else. So many I forget most of them, yeah.

[Q]MR: Did you ever cross paths with Joe Williams?

[A]GW: Oh, I see Joe Williams all the time. Every day. His piano player doesn't want to make the trip out to my trio, yeah. for me.

[Q]MR: What's it like working with Joe Williams?


[A]GW: Easiest job on earth. But I don't have to work with his piano player, they have all this stuff set up. But when I worry about it. " The Blues. We do the Blues in any ... all kind of keys. We do a ballad now and then, but it's very easy working for him.

[Q]MR: That's interesting. You know the non-jazz musician always finds it fascinating when if you watch someone and they say B flat, and that's all the band needs. You know, that's one great thing about ...

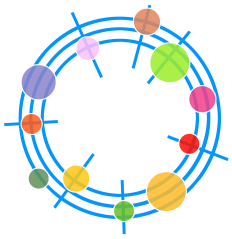
[A]GW: Well you're supposed to know the tunes, you know?

[Q]MR: That's another thing -- for this type of festival, too, when you watch the bands up there, and you've really got to have a repertory in your head of tunes. Because some of them aren't groups that work together all

Apply this choice to all 'Joe'

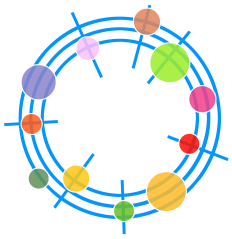
 Joe Williams

Ignore this partial



Transcript Analyzer

- An interface to curate the transcripts and verify detected names.
- Implements off the shelf NLP (NLTK) to attempt to locate additional names not in our directory as well as corporate names and locations.
- Global rule system, as we process more transcripts the system is being trained.
- Using URIs to represent entities we can quickly see where we are discovering new material.
 - 50 Transcripts
 - 1800 person entities tagged.
 - 250 names tagged without authoritative URI.
 - Knowledge Creation



New Dataset

- We have created a new LOD dataset now of jazz musician's relationships.
- Our next steps are:
 - Visualize.
 - Further qualify the rel:knowsOf relationships.
 - Provide access to the data created.



Visualize

LINKED JAZZ



Fixed



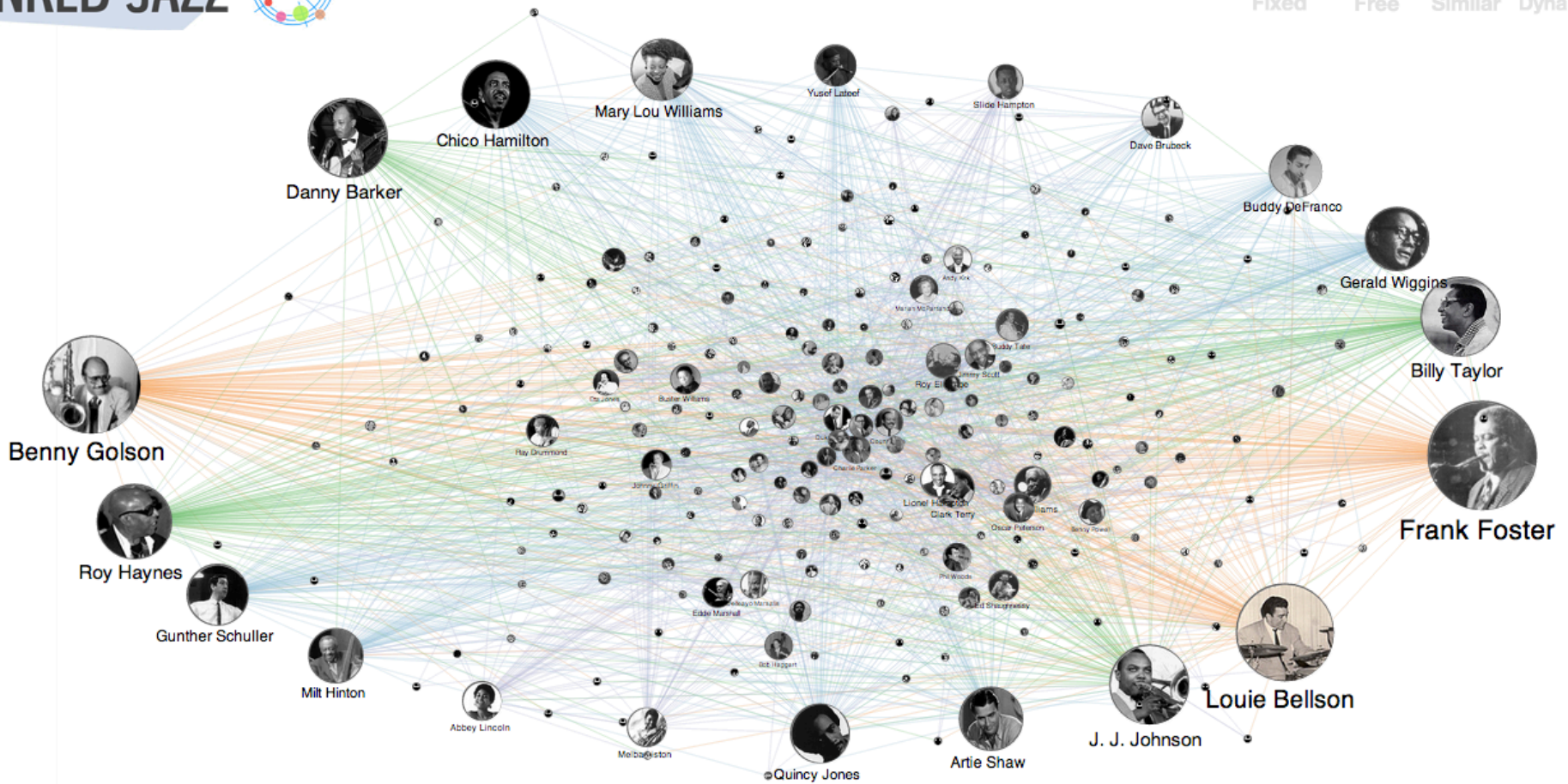
Free



Similar



Dynamic



<http://linkedjazz.org/network/>




Qualify Relationships – 52nd St.

- Recruit jazz experts and enthusiasts to help categorize relationships based on transcript text.
- We use existing vocabularies to build the data set: Foaf, Relationship Vocabulary, Music Ontology
- The interface is critical for crowdsourcing tools, we work with user experience experts and conduct user studies to refine our public facing tools



Qualify Relationships – 52nd St.

LINKED JAZZ 

Mary Lou Williams talks about **Cab Calloway** (1 of 1 mentions) [Transcript Source](#) [About](#) | [Account](#)

Howard Burley
Fletcher Burley
Virginia Burley
Don Byas
Blanche Calloway
Cab Calloway
Thelma Carpenter
Benny Carter
Doc Cheatham
Bill Clark
Kenny Clarke

Interviewer

When did you first run into Lester Young?

When Ben Webster left [the] Andy Kirk to join **Cab Calloway**, Lester Young joined us in Kansas City, but he didn't stay with us very long because shortly after that he went to Count Basie's band.

Based on this text, how would you describe this relationship?

Knows of
Has met
Acquaintance of
Close friend of
Collaborated with
Influenced by
Mentor of
Skip
Add Comment

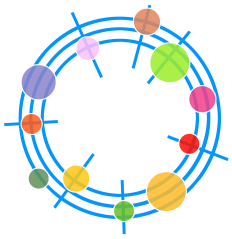
5% Complete!

What is this? Feedback



Provide Access

- We provide a SPARQL endpoint.
- But also a traditional API:
 - <http://linkedjazz.org/api/>
 - Can return:
 - JSON
 - N-Triples
 - Gephi graph files (GXEF)

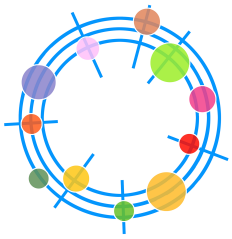


Learn and Grow as a Team

- Experience through doing.
- Empower graduate students with skills and practical experience working with a LOD project.
- Use the project as a vehicle to make intra- and inter-intuitional collaborations.



Linked Jazz Team April 2013



Next Steps

- Refactor our prototype tools into sustainable open source projects.
- Redesign 52nd St. based on user study groups.
- Work on emerging collaborations with Jazz Centers.



<http://www.linkedjazz.org>