

## Jazz History as Linked Data: The *Linking Lost Jazz Shrines* Project

Zakiya Collier, Sarah Ann Adams, M. Cristina Pattuelli, Obden Mondésir

[Data for History 2020: Modelling Time, Places, Agents](#)

Linking Lost Jazz Shrines (LLJS) is a project developed through a collaboration between the Weeksville Heritage Center (<https://www.weeksvillesociety.org/>) and the Semantic Lab at Pratt Institute (<http://semlab.io>) to investigate the application of linked open data technologies to the Weeksville Lost Jazz Shrines of Brooklyn (WLJSB) oral history collection (<http://bit.ly/WLJSBFindingAid>). WLJSB is a rich collection of oral histories that documents Central Brooklyn's cultural legacy of jazz between the 1930s and the 1960s. The primary aim of the project is to make the WLJSB archival collection more discoverable and accessible to researchers and the general public alike by leveraging linked data principles and technologies. At the same time, the project will enrich and expand the existing network of jazz musicians developed by the [Semantic Lab's](#) Linked Jazz Project (<https://linkedjazz.org/>) with an infusion of influential, yet lesser-known African-American artists and venues. Finally, the project has the potential to serve as a roadmap for similar projects aimed at transforming textual archival documents and other materials into linked data.

The paper describes the methods used to generate linked open data from text using a dedicated tool, recently developed by the Semantic Lab (Sélavy, available from <http://tools.semlab.io/>) to support Name Entity Recognition and Linking. It also discusses the integration of the data generated from the WLSJB collection into the Linked Jazz knowledge graph, both at the ontology and the instance level, to enable more complex queries and foster new lines of research. While Linked Jazz provides the representation of personal and professional relationships between jazz musicians (Semantic Lab, 2020), LLJS would contribute further semantic extension and refinement of the relations in order to represent not only individual musicians, but also music groups and the jazz venues where all the musicians performed, the “lost jazz shrines” of Brooklyn. The paper also describes our experience using Wikibase, the technological infrastructure that enables data to be modeled, stored, managed and queried in a unified data space. Our local Wikibase instance (<https://base.semlab.io/>) provides the collaborative platform needed for the development of the project, facilitating the expansion of the data model as well as the contextualization of the data, and ultimately allowing for the consumption of the data through visualizations and queries.

**Key Words:** Oral Histories; Linked Open Data; Semantic Technologies; Archives; Wikibase; Collections as Data; Jazz History; Brooklyn

*The initial phase of Linking Lost Jazz Shrines was conducted from January 2019 through August 2020 with generous support from the Andrew W. Mellon Foundation “Collections as Data: Part to Whole” grant. The current phase of the LLJS Project is generously supported by the Metropolitan New York Library Council (METRO) “Equity in Action” grant.*