introduction to digital scholarship

Brandon Locke
@brandontlocke | brandontlocke.com

about me

- Involved in digital history/humanities/scholarship since 2010
- 2010-2013 MA History, University of Nebraska-Lincoln
  - Produced a digital MA thesis, received Certificate in DH, Project Manager for History Harvest, RA on several projects
- 2013-2014 MSLIS University of Illinois at Urbana-Champaign
  - RA on several digital projects
- 2014-2019 Director, Lab for the Education and Advancement in Digital Research (LEADR), Michigan State University
  - Student-focused lab in History and Anthropology Departments
  - Developed digital curriculum, supported student work, taught technical workshops
  - Worked with graduate students on research projects
  - Collaborated with cross-campus partners on digital scholarship curriculum, programming, projects

roadmap for this afternoon

- definitions
- values
- people

<<break>>

- teaching
- project development & tech

definitions

defining “digital scholarship”

- The “**use of digital evidence and method, digital authoring, digital publishing, digital curation and preservation, and digital use and reuse of scholarship.**” - Abby Smith Rumsey in Mulligan, Rikk. Supporting Digital Scholarship. SPEC Kit 350. ARL, May 2016.
- See Jason Heppler’s https://whatisdigitalhumanities.com/
- Definitions are fluid, highly contextual, and may depend significantly on the staff, services, and equipment that are available in a particular location
- Definitions can gatekeep, or they may be generative

affordances of digital objects

procedural	**extent to which computer can exert processing power on objects to extend and/or enable a question**, e.x. Counting word frequency, finding names and places in text, finding patterns in numismatic data
spatial	**extent to which objects have a spatial component** that can be leveraged, e.x. place names can be geocoded and mapped
encyclopedic	**extent to which objects can be made more comprehensively accessible**, e.x. searches using metadata and plain text, aggregation by theme, space, author information
participatory	**extent to which the objects can be made available in a manner that invites interaction**, e.x. data sharing, crowdsourced transcription

slide adapted from Thomas Padilla; concept adapted from Janet H. Murray, *Design Exercise: Affordance Grid*, 8.14.2012

what digital scholarship is not...

- analysis of big data (exclusively)

what digital scholarship is not...

- analysis of big data (exclusively)
- something only done by computer/data scientists (or even ‘tech savvy’ people)

what digital scholarship is not...

- analysis of big data (exclusively)
- something only done by computer/data scientists (or even ‘tech savvy’ people)
- something you can do all by yourself

values

values of digital scholarship

- openness
- collaboration
- collegiality and connectedness
- diversity
- experimentation

From Lisa Spiro, “This is why we fight: Defining the values of digital humanities,” Debates in the Digital Humanities, 2012 & Kristen Mapes “Teaching Values, Not Definitions: Experiences and Research in the Introductory Digital Humanities Course” kristenmapes.com/siue2019

openness

- open ideas, open access, open source software, transparency [Spiro]
- promotes interoperability, discovery, usability, and reusability (Christine Borgman) & democratize knowledge to reach out to ‘publics,’ share academic discoveries, and invite an array of audiences to participate in knowledge production” (Draxler et al.). [via Spiro]

collaboration

- free flow of information between a mixture of people can lead to breakthroughs (Stephen Johnson via Spiro)
- DS needs many different sets of expertise to succeed
- collaboration also often includes community partners

collegiality and connectedness

- open to people with different skills, professional backgrounds, visions
- open to non-hierarchical, non-disciplinary, inter-professional work (Spiro)
- ethical labor and mentoring
    - For student labor, see Miriam Posner et al., “A Student Collaborators’ Bill of Rights,” *UCLA Digital Humanities* & Spencer Keralis, “Disrupting Labor in Digital Humanities; or, The Classroom is Not Your Crowd” in Disrupting the Digital Humanities
- ethical treatment of source materials & affected communities

diversity

- the community is more vibrant, discussions are richer, and projects are stronger if multiple perspectives are represented (Spiro)
- race/gender/sexuality/nationality/ability/age/disciplinary diversity in the community, as well as in the content of the projects
- equity and justice also tie into this

experimentation

- openness & support for risk taking, entrepreneurship, innovation (Spiro)
- leveraging information technology to explore data, digital humanities casts intellectual problems as experiments: What is the effect of modeling the data in a particular way? What happens when we visualize data or use text mining tools to discover patterns in it? (Svensson, “The Landscape of Digital Humanities” via Spiro)
- recognize the value of failure and the ability to work through it (Spiro)

select a project from the following list

bit.ly/dsiprojectlist

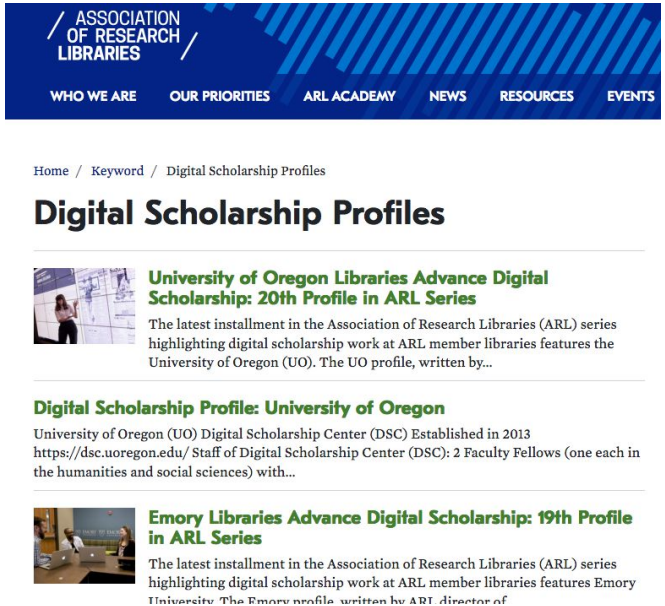
and evaluate them according to the first page of the following worksheet

bit.ly/dsiprojecteval

people

digital scholarship models

Mulligan, Rikk. Supporting Digital Scholarship. SPEC Kit 350. Washington, DC: Association of Research Libraries, May 2016. doi.org/10.29242/spec.350



ARL Digital Scholarship Profiles arl.org/arl-terms/digital-scholarship/

Greenhall, M. (2019). Digital scholarship and the role of the Research Library, RLUK Report. rluk.ac.uk/digital-scholarship-and-the-role-of-the-research-library-an-rluk-report

ds support

includes organizational (project planning), technical (database development), and methodological (computational text analysis)

Right: Survey results from SPEC 350 (2016)

1. Please indicate where a researcher at your institution (whether faculty, student, or other researcher) can find support for the digital scholarship activities listed below. Check all that apply
N=73

Digital Scholarship Activities	In the library	Elsewhere in the institution	Elsewhere outside the institution	N
GIS and digital mapping	65	45	9	72
Digitization/imaging of analog material	71	22	9	71
Making digital collections	67	20	11	71
Metadata creation	67	14	7	70
Digital preservation	69	13	8	69
Data curation and management	65	21	9	69
3-D modeling and printing	42	59	10	69
Statistical analysis/support	40	57	9	69
Digital exhibits	66	23	9	67
Project planning	61	40	11	67
Digital publishing	62	25	12	67
Project management	49	39	10	61
Computational text analysis/support	49	36	7	61
Interface design and/or usability	48	31	11	61
Visualization	49	41	7	60
Database development	40	41	9	58
Technical upkeep	45	38	8	56
Encoding content (e.g., TEI markup)	44	20	10	52
Developing digital scholarship software	35	28	16	51
Other DS activity	19	10	5	20
Total Responses	73	68	29	73

ds activities

- workshops
- embedded teaching
- IL sessions
- consultations
- project partnerships
- grant writing
- space & infrastructure development
- digitization and cataloging/metadata
- collection development
- speaker series

common digital scholarship models

DS most often operates in a “mixed economy,” where work is done in many different areas of campus; even when the library is not at the center of the work, it is often a broker between different spaces

- Centralized (dedicated staff; often organized around a lab/center)
- Decentralized (distributed expertise)
- Hubs (multiple clusters/spaces in library and around campus)
- Single person (one dedicated staff/admin person)

centralized model

- a unit or department organized around digital scholarship support
- often includes a lab or a center
- can be a joint effort between a library and a department/college

hub model

- multiple units/spaces supporting digital scholarship in different ways
- often includes library and non-library hubs
- hubs are often in communication with each other, but not always coordinated by a central group
- trend towards this model in recent years (SPEC 350, p6)

decentralized model

- no formal organization around DS
- support is offered by domain experts (e.g. special collections, metadata librarian, GIS librarian)

single person model

- one dedicated specialist who may coordinate other librarians
- often seen as a pilot or transition step
- have been dubbed “miracle workers”

Has your institution discussed reorganization or the development of service points?

What has driven these conversations? Are there any portions of these conversations that have left you particularly concerned or optimistic?

break

teaching

course-based teaching

- students don't care about "Digital Humanities" or "Digital Scholarship" (Cordell. *How Not to Teach Digital Humanities* in Debates in the Digital Humanities 2016.)
- digital scholarship gives librarians an opportunity to delve deeper into information, digital, data, and algorithmic literacies (Locke. Digital Humanities Pedagogy as Essential Liberal Education. Digital Humanities Quarterly 2017.)
    - Information literacy: ability to dive deeper into finding & evaluation sources, metadata, contextual information
    - Digital literacy: understand affordances of digital reading; learn to communicate digitally
    - Data literacy: help students better navigate a data-driven world (Fonticharo & Oehrli datalit.sites.uofmhosting.net) & integrate research data management (Carlson & Johnston datainfolit.org)
    - Algorithmic literacy: better understand the constructed nature of algorithms and their impact on research and society

teaching challenges

- assignment is well-incorporated into classes
  - student work makes sense within the context of the class
  - instruction isn't too early or too late in the project
  - faculty are actively supportive
- experimentation is rewarded and a safety net is present
  - grades feel like a high-stakes game for students
- value & purpose of digital project is made clear
  - what does this technology do for their understanding of the subject? For their everyday lives?
- goals, expectations, roles are clearly defined with faculty members
  - Not sending mixed messages, focusing on different things, etc
- students are creating their own projects; not faculty projects
  - UCLA Student Collaborators' Bill of Rights & Keralis "Disrupting Labor in Digital Humanities"

teaching speedbumps

- communication with faculty
- never undermine student challenges
- not every student has a machine with an operating system
- software dependencies are weird and everyone will have a different problem
- students miss classes
- data management is hard (and most people haven’t been taught!)
- movable seating, markerboards, pens and paper are effective technology
- rubrics are good, but can be really limiting

projects

project challenges

- sustainability
  - funding
  - technological
  - personnel
- project management
  - communication
  - scoping
- labor & time management
- trust & permission from admin
  - failure
  - content
  - SERVERS!

project development

Miriam Posner, UCLA

take a project, dig into the documentation and the functionality, break it into its component parts, and begin to understand how it was built

Concept/slides from Posner, Miriam.
How Did They Make That? (2014) archive.org/details/howdidtheymakethat

sources, processed and presented

break the project down into three primary components: the original sources, the process of turning those sources into some kind of data, and the technology to take that data and present it on the web

each of these steps includes many scholarly decisions that have an impact on the project as a whole

Concept/slides from Posner, Miriam. How Did They Make That? (2014)
archive.org/details/howdidtheymakethat

sources, processed and presented

- documents
- articles
- letters
- photographs
- artifacts
- text
- numbers
- video
- sound

Concept/slides from Posner, Miriam. How Did They Make That? (2014)
archive.org/details/howdidtheymakethat

sources, processed and presented

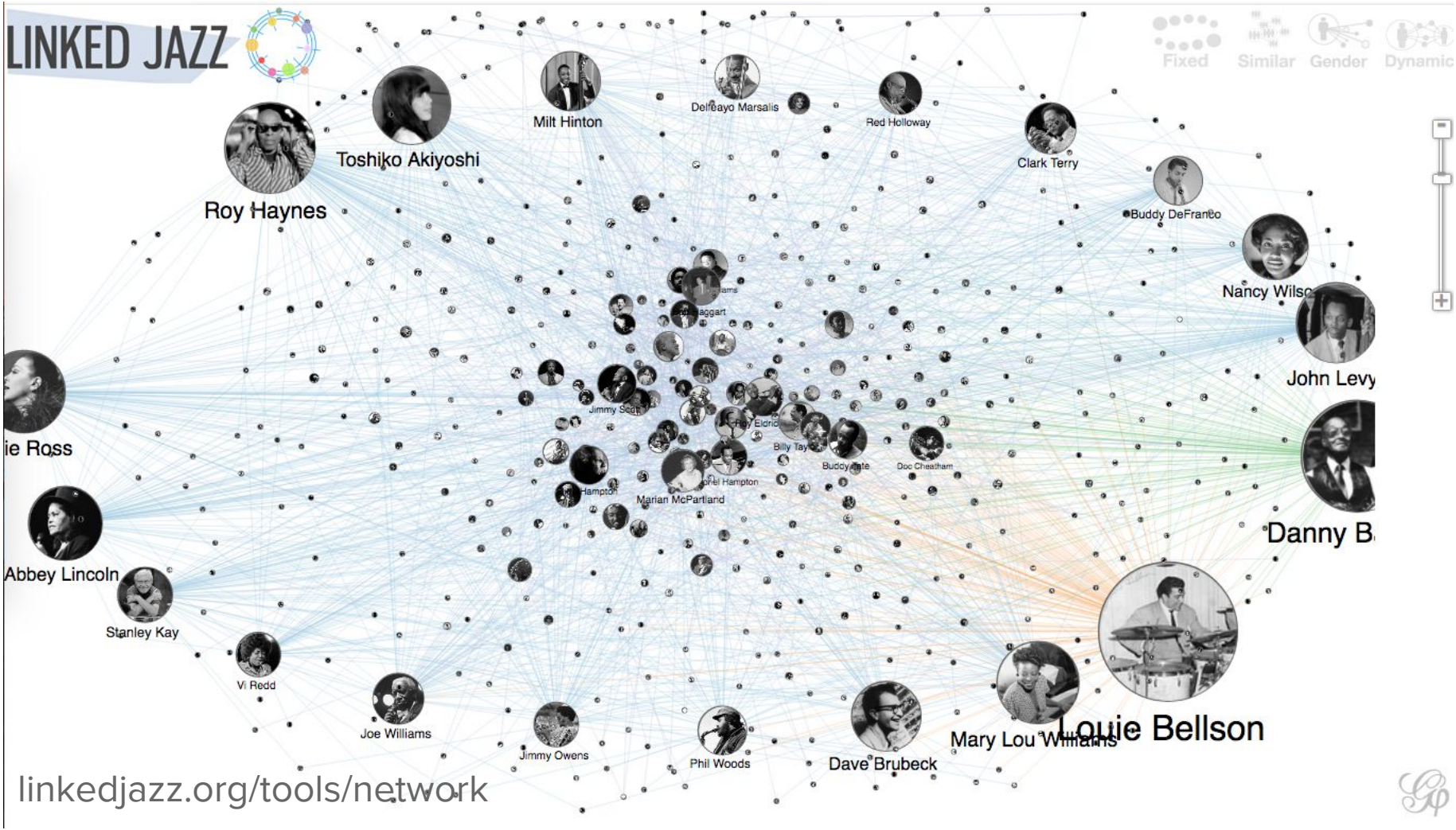
- digitized
- organized
- edited
- annotated
- extrapolated
- photographed
- enhanced
- OCR’d
- quantified

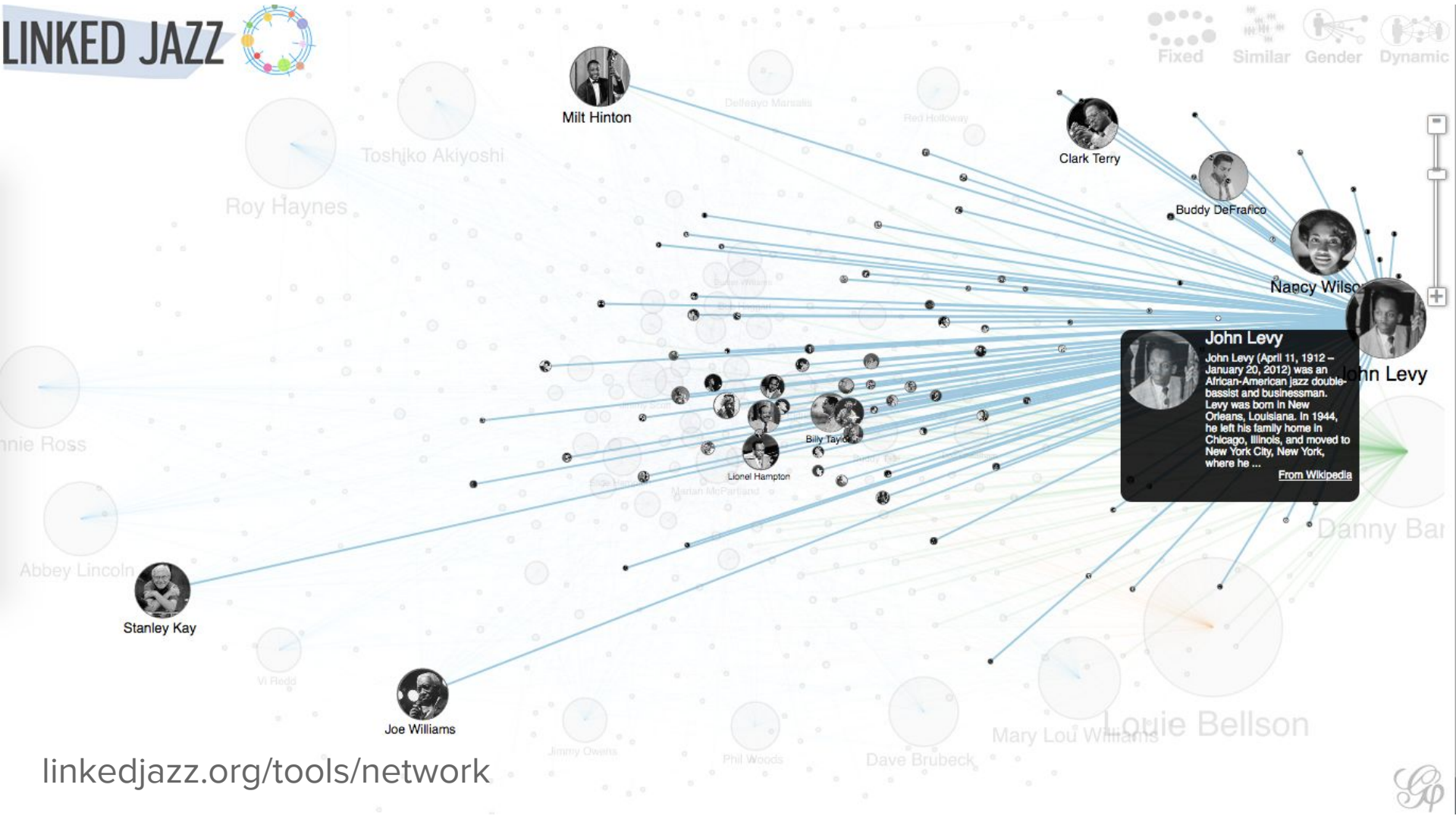
Concept/slides from Posner, Miriam. How Did They Make That? (2014)
archive.org/details/howdidtheymakethat

sources, processed and presented

- visualized
- made web accessible
- mapped
- made searchable or queryable
- made interactive

Concept/slides from Posner, Miriam. How Did They Make That? (2014)
archive.org/details/howdidtheymakethat


LINKED JAZZ
Fixed
Similar
Gender
Dynamic
Milt Hinton
Delfeayo Marsalis
Red Holloway
Toshiko Akiyoshi
Roy Haynes
Clark Terry
Buddy DeFranco
Nancy Wilso
John Levy
Jimmy Scott
Billy Tay
Buddy Tate
Doc Cheatham
Hampton
Marian McPartland
Danny B
Abbey Lincoln
Stanley Kay
Vi Redd
Joe Williams
Jimmy Owens
Phil Woods
Dave Brubeck
Mary Lou Williams
Louie Bellson
linkedjazz.org/tools/network



linkedjazz.org/tools/network

sources, processed and presented

Oral History Transcripts

We've utilized over 50 transcripts of oral history interviews in our project. The original interviews come from the Hamilton College Jazz Archive, Rutgers Institute for Jazz Studies Archives, Smithsonian Jazz Oral Histories, UCLA's Central Avenue Sounds Series, and the University of Michigan's Nathaniel C. Standifer Video Archive of Oral History.

DBPedia

The seed of our dataset is a list of names extracted from DBPedia, the Linked Open Data version of Wikipedia. At the beginning of our project, we created a starter list of names of jazz musicians. This list, comprised of **9300 names**, was generated by **filtering the DBpedia RDF extracts for Jazz related individuals**.

This list of URIs was used to match against the names our transcript analyzer recognized using Natural Language Processing, so we could assign them URIs.

linkedjazz.org/data-sources

sources, processed and presented

To create LOD, we developed a suite of tools: a transcript analyzer, a name mapping and curator tool, and a crowdsourcing tool. These tools operate together to find names mentioned during the interview in order to assign a positive identification to each, disambiguating names using online resources like DBpedia and VIAF. The transcript analyzer also recognizes the question and answer structure of the oral history. As people are mentioned by an interviewee, simple RDF triples between interviewee and persons mentioned are created in the form of *knowsOf*. These triples can then be mapped to the correlating block. For more on the transcript analyzer and Name Mapping tool, see our Tools page.

In the next step, the interview question and answer blocks are passed to Linked Jazz 52nd Street, our crowdsourcing tool. Volunteers are presented with these snippets of interview text and asked to assign more granular terms to describe the relationship between the interviewee and the person mentioned. These relationships include:

has met
is an acquaintance of
is a friend of
is a close friend of
is influenced by
is a mentor of
collaborated with:
- **was in a band together with**
- **played with**
- **was a member of the band of**
- **toured with**
- **was the bandleader for**

linkedjazz.org/about-the-project/

sources, processed and presented

Relationship Data

When fed with an interview transcript, the Analyzer breaks the text into question-and-answer blocks, and leverages this structure to automatically identify a connection between two individuals based on the assumption that if the interviewee mentions a person, she or he must at least *know of* the person cited. An RDF triple is then generated that states this basic relationship:

```
<http://linkedjazz.org/resource/Clark_Terrry> <http://purl.org/vocab/relationship/knowsOf>
```

In addition to automatic processing, we also employ *human* processing to generate data. Through our crowdsourcing tool 52nd Street, volunteers are presented with snippets of interview text and asked to assign more specific terms to describe the relationship between the interviewee and the person mentioned. When the volunteer chooses a relationship from a list of options, this action generates a triple that will be stored in our dataset.

linkedjazz.org/data-productionworkflow-draft

sources, processed and presented

Transcript Analyzer

The Linked Jazz Transcript Analyzer structures digital archival documents for different purposes and identifies named entities in texts. In the context of the project, we use the Analyzer to upload interview transcripts from open access archives and to identify personal names cited in interview transcripts by leveraging the above-mentioned Linked Jazz Name Directory. The analyzer also employs natural language processing to locate names that are not present in the directory. In these instances, we relate the newly found names to URIs from the name authority files, or, if the name is not found in the authorities, we mint new URIs that we then host on the Linked Jazz namespace. Finally, the analyzer tool breaks interview transcripts down into discrete segments of questions and answers, which are later employed in the Linked Jazz 52nd Street tool.

linkedjazz.org/tools/transcript-analyzer

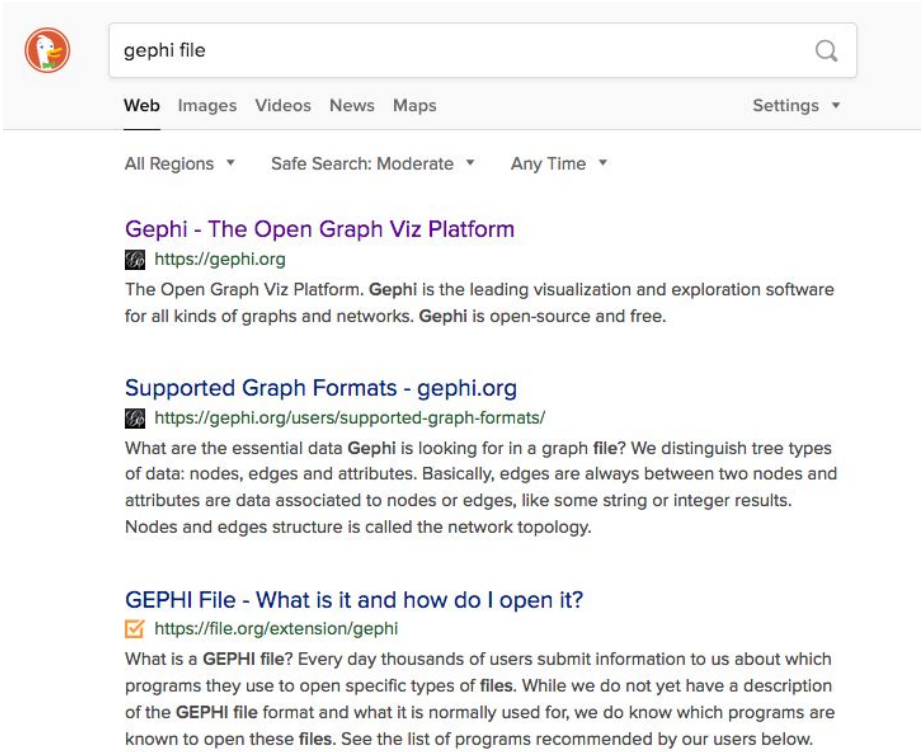
sources, processed and presented

Linked Jazz 52nd Street

Linked Jazz 52nd Street is a crowdsourcing tool that allows jazz experts and enthusiasts to assist us in deciding what type of relationship two individuals share based on interview transcripts. While we can assume that jazz artists who cite other jazz artists in their interviews have some kind of association with them, this relationship could be anything from close friendship and collaboration to just knowing the other person exists. Linked Jazz 52nd Street addresses this problem. This tool is a web-based application that asks contributors to classify the relationship between two jazz artists according to a menu of options. This assessment is facilitated by presenting the contributor with interview excerpts referencing the individuals in question. Results are tallied and converted into RDF statements that feed the project's LOD dataset.

linkedjazz.org/tools/linked-jazz-52nd-street

sources, processed and presented



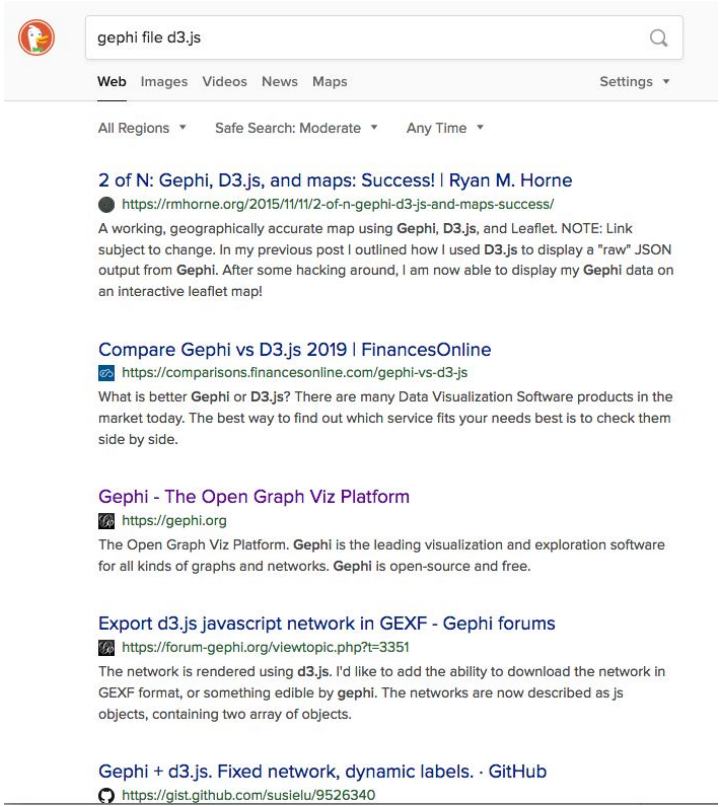
linkedjazz.org/tools/network

sources, processed and presented

```
18
19
20
21 <script src="https://ajax.googleapis.com/ajax/libs/jquery/1.7.2/jquery.min.js"></script>
22 <script src="js/jquery.history.js"></script>
23 <script src="js/jquery.rdfquery.core.min-1.0.js"></script>
24 <script src="js/dragdealer.js"></script>
25 <script src="js/d3.v2.min.js"></script>
26 <script src="js/vex.min.js"></script>
27 <script src="js/network.js"></script>
28
```

View source: linkedjazz.org/tools/network

sources, processed and presented



how did they make linked jazz?

1. collected interview transcripts
2. extracted (manually and automatically) names and connected them together in linked open data
3. used that linked open data in Gephi to construct networks
4. published that network using d3.js and other js packages

Read through each of Posner's linked *How Did They Make That?* posts.

Select one of the projects and dig into it a bit deeper

miriamposner.com/blog/how-did-they-make-that

select a project from the following list

bit.ly/dsiprojectlist

and evaluate them according to the first page of the following worksheet

bit.ly/dsiprojecteval

thank you!

Brandon Locke
@brandontlocke | brandontlocke.com