Visualizing MARC Data in a Graph Database:
American Show Tunes 1940-1945

University of Maryland iSchool Capstone Project Design & Research: Tracy Meehleib
University of Maryland iSchool Capstone Project Advisors: Dr. Michael Kurtz & Dr. Richard Marciano
Library of Congress Capstone Project Advisors: Ardie Bausenbach & Morgan Cundiff
Library of Congress CMDA Program Support: Beacher Wiggins & Sally McCallum
Visualizing MARC Data in a Graph Database:
American Show Tunes 1940-1945

Research Questions

1) What are the challenges involved in identifying and appropriate, relevant and viable set of MARC records in a large research library, and once identified, how feasible and useful is it to import them into a graph database?

2) Is a graph database—combined with graph visualizations—useful for bibliographic search and discovery?

3) How would MARC data be modeled in a graph database?

4) What are the MARC to graph data model conversion issues and options?

5) Can a standard workflow be developed that can be applied to convert any MARC record set into a graph database?

6) Can an open source toolset be identified and used to accomplish this work?

7) What are graph database visualization issues and options?

8) What are the potential benefits of converting MARC data to a graph data model?
Visualizing MARC Data in a Graph Database:

American Show Tunes 1940-1945

Overview of Capstone Project Design, Research, and Development Methodology

1) Research graph theory, Neo4j documentation, visualization tools, potential tool set
2) Identify and acquire tool set required for project (hardware and software)
3) Identify and acquire record set (53 test records—then 3,694 records of 67,000 for actual capstone project)
4) Analyze and identify relevant MARC fields/subfields (MarcReport)
5) Identify and test MARC conversion process, e.g., .XML > .MRC > .CSV (MarcEdit)
6) Model data — identify relevant nodes, relationships, and properties (Pencil & Paper > Arrow Tool)
7) Map MARC fields and subfields to data model
8) Parse and extract MARC data to conform to data model — and convert from .XML > .MRC > .CSV (MarcEdit)
9) Load MARC-based .CSV data into graph database and data model
10) Refine data model as needed and retest
11) Analyze graph data using queries and visualizations
12) Draw conclusions
13) Identify next steps
Visualizing MARC Data in a Graph Database:
American Show Tunes 1940-1945

Record/Data Conversion Process

1) Identify relevant MARC fields and subfields to extract and model
2) Convert data (relevant fields and subfields) from .XML > .MRC > .CSV using MarcEdit
3) Load MARC-based .CSV data into graph database and data model

```sql
1 // persons P1
2 USING PERIODIC COMMIT
3 LOAD CSV WITH HEADERS FROM "file:///Users/tracymeeleib/Desktop/m1508final.csv" AS row
4 MERGE (p:Person {p1ID: row.P1ID, p1Date: row.P1Date, p1Role: row.P1Role});
```
Visualizing MARC Data in a Graph Database:
American Show Tunes 1940-1945

Data Modeling (in Arrow) and Testing (in Neo4j)

1) Model Data (Pencil & Paper > Arrow Tool)

2) Merge data to conform to data model in Neo4j
Visualizing MARC Data in a Graph Database:

American Show Tunes 1940-1945

Analysis & Results

Macro View of Data (Collaborative Networks, Nodes, and Relationships)

Micro View of Data (Nodes and Relationships)
Visualizing MARC Data in a Graph Database:

American Show Tunes 1940-1945

Potential Implications and Possibilities for Search and Discovery

Collaborative Networks Analysis and Linked Data Possibilities
Visualizing MARC Data in a Graph Database:

American Show Tunes 1940-1945

Potential Implications and Possibilities for Search and Discovery

Collaborative Networks Analysis and Linked Data Possibilities: Sammy Cahn
Visualizing MARC Data in a Graph Database:
American Show Tunes 1940-1945

Library of Congress Music Division — Popular Use Case — Filtering a Query

1) What songs were in these shows?
2) What shows had these songs in them?
Visualizing MARC Data in a Graph Database:

American Show Tunes 1940-1945

Next Steps

1) Document basic workflow, hardware requirements, and toolset for reuse

2) Present project results within the Library of Congress

3) Publicize GitHub Graph Gist when ready < https://gist.github.com/tmeehleib/3c0ced8a33887294544cd71880b15bf5 >

4) Explore visualization technologies that further leverage Neo4j graph data (Tableau, Linkurious, Ogma, etc.)

5) Experiment with other MARC data sets (Music, Prints & Photographs, Kluge Center, Digital Lab, etc.)

6) Experiment with adding linked data to graph data

7) Explore using graph data to model current MARC > BIBFRAME 2.0 mapping for NDMSO

8) Consider applying for a Kluge Scholarship to continue and expand work in this area
Visualizing MARC Data in a Graph Database: American Show Tunes 1940-1945

THANK YOU!!!