

Enhancing Access to Legacy of Slavery Records using Generative AI

Rajesh Kumar Gnanasekaran

Lori Perine

Richard Marciano

Mark Conrad



Background on MSA's LoS

Title: Why focus on Maryland State Archives' Legacy of Slavery (LoS)?

- The *Maryland State Archives (MSA)* undertook a *multi-year* effort with about 100 staff since 2001 to digitize historical records from the *LoS* project. Project to discover stories of unknown 'heroes' of slave flight and resistance.
- This effort resulted in a large, digital database containing over *400,000* records, including Certificates of Freedom (CoF), Domestic Traffic Ads (DTA), and Manumissions.
- Despite this significant achievement, the actual usage and engagement with the dataset remain low.



Leveraging Generative AI & Knowledge Graphs

“ChatLoS” using *Generative AI & Knowledge Graph-RAG Framework*

- Chatbot (“ChatLoS”) that pulls context from the *Domestic Traffic Ads (DTA)* dataset and provides *context-aware* responses
- Demonstrates *interactive natural language* conversational capabilities
- *Knowledge Graphs (KGs)* serve the purpose of getting the data “AI-Ready” by providing structured source of truth context (through nodes, relationships)
- Study included three datasets from LoS together (*CoF, DTA, Manumissions*) for cross-dataset insights.
- Also includes *a systematic user evaluation grounded on Activity Theory* by involving *two separate IRB approved* user studies with well experienced MSA’s experts



UNIVERSITY OF
MARYLAND

CLARKE
FORWARD

Prior solutions only looked at using one dataset in a context-aware manner and with non-AI ready data



Leveraging Generative AI & Knowledge Graphs

“ChatLoS” using *Generative AI & Knowledge Graph-RAG Framework*

- Also includes *a systematic user evaluation grounded on Activity Theory* by involving *two separate IRB approved* user studies with well experienced MSA’s experts
- Prior solutions only looked at using one dataset in a context-aware manner and with non-AI ready data
- User study findings were *positive with room for improvement*.



How KG-RAG works with ChatLoS?

- *Knowledge Graphs (KGs)* provide structured source of truth context (through nodes, relationships)
 - Nodes and relationships built by connecting multiple datasets together by integrating *CoF*, *DTA*, *Manumissions* for cross-dataset insights.
- KG-RAG connects the KG and GPT to create GraphQL or Cypher Queries dynamically and provide that to the GPT to generate natural language responses, similar to simple-RAG but more powerful
- Used Neo4J desktop to import datasets and create nodes and relationships



Node Label
EnslavedPerson
SlaveOwner
ManumissionRecord
CertificateOfFreedom
DomesticTrafficAd

Relationship Type	Source Node	Target Node
OWNED_ENSLAVED_PERSON	SlaveOwner	EnslavedPerson
PLACED_DOMESTIC_TRAFFIC_AD	SlaveOwner	DomesticTrafficAd
LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN	EnslavedPerson	DomesticTrafficAd
SIGNED_MANUMISSION	SlaveOwner	ManumissionRecord
MANUMITTED_THROUGH	EnslavedPerson	ManumissionRecord
WAS_GRANTED_FREEDOM_THROUGH	EnslavedPerson	CertificateOfFreedom
WHO_ENSLAVED_WAS_FREEDOM_FROM	CertificateOfFreedom	SlaveOwner

Database Information

Use database

neo4j

Node labels

*(61,291)

CertificateOfFreedom

DomesticTrafficAd

EnslavedPerson

ManumissionRecord

SlaveOwner

Relationship types

*(60,482)

LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN

MANUMITTED_THROUGH

OWNED_ENSLAVED_PERSON

PLACED_DOMESTIC_TRAFFIC_AD

SIGNED_MANUMISSION

WAS_GRANTED_FREEDOM_THROUGH

WHO_ENSLAVED_WAS_FREEDOM_FROM

neo4j\$

```

1 MATCH (so:SlaveOwner)-[:SIGNED_MANUMISSION]->(mr:ManumissionRecord),
2   (so)-[:PLACED_DOMESTIC_TRAFFIC_AD]->(dta:DomesticTrafficAd),
3   (ep:EnslavedPerson)-[:LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN]->(dta)
4 RETURN so, mr, dta, ep;

```

Graph

Table

Text

Code

Cross-dataset querying with DTA, Manumissions, and Certificates of Freedom



How many enslaved individuals were listed for sale in the Dorchester County in the State of Maryland between 1825 and 1830?



Based on the data available, there were 78 enslaved individuals listed for sale in Dorchester County, Maryland, between the years 1825 and 1830.

Generative AI explanation behind the response

What the AI did:

The AI interpreted your question: 'How many enslaved individuals were listed for sale in the Dorchester County in the State of Maryland between 1825 and 1830?', generated a Cypher query to explore graph relationships in the Knowledge Graph, executed it against Neo4j, and composed an answer from the query results.

Internal Logs (Advanced):

> Entering new GraphCypherQAChain chain...

Generated Cypher:

```
MATCH (e:EnslavedPerson)-[:LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN]->(ad:Do
WHERE ad.county_name = 'Dorchester' AND ad.Sale_Date >= '1825' AND ad.S
RETURN COUNT(DISTINCT e) AS Number_of_Enslaved_Persons_Listed_For_Sale
```

Full Context:

```
[{'Number_of_Enslaved_Persons_Listed_For_Sale': 78}]
```

> Finished chain.


DEBUG: Chain Result: {'query': 'How many enslaved individuals were list


Clear Chat


Figure 37: ChatLoS v3's answer showing “explainable” response for an aggregate multi-attribute filter question



ChatLoS enhanced with KG-RAG using Neo4j and Generative AI—Cross-dataset querying with DTA, Manumissions, and Certificates of Freedom

 How many certificates of freedom are there?

 There are a total of 23,638 Certificates of Freedom in the dataset.

 Generative AI explanation behind the response

What the AI did:

The AI interpreted your question: 'How many certificates of freedom are there?', generated a Cypher query to explore graph relationships in the Knowledge Graph, executed it against Neo4j, and composed an answer from the query results.


Internal Logs (Advanced):

```
> Entering new GraphCypherQAChain chain...
Generated Cypher:


MATCH (c:CertificateOfFreedom)
RETURN COUNT(c) AS TotalCertificatesOfFreedom


Full Context:
[['TotalCertificatesOfFreedom': 23638]]


> Finished chain.
DEBUG: Chain Result: {'query': 'How many certificates of freedom are there?'}
```


 Clear Chat


Ask a question (e.g., 'Can you trace the life of an individual who has appeared in the Domestic


 How many certificates of freedom are there?

 There are a total of 23,638 Certificates of Freedom in the dataset.

 Generative AI explanation behind the response

 How many manumissions are there?

 There are a total of 7,235 manumission records in the database.

 Generative AI explanation behind the response

What the AI did:


The AI interpreted your question: 'How many manumissions are there?', generated a Cypher query to explore graph relationships in the Knowledge Graph, executed it against Neo4j, and composed an answer from the query results.

Internal Logs (Advanced):

```
> Entering new GraphCypherQAChain chain...
Generated Cypher:
cypher
MATCH (m:ManumissionRecord)
RETURN COUNT(m) as manumissions_count

Full Context:
[['manumissions_count': 7235]]

> Finished chain.
DEBUG: Chain Result: {'query': 'How many manumissions are there?', 'res'}
```

 Clear Chat

Ask a question (e.g., 'Can you trace the life of an individual who has appeared in the Domestic

Figure 38, 39: ChatLoS v3's KG RAG-based answers responding for multi-dataset aggregate questions



ChatLoS enhanced with KG-RAG using Neo4j and Generative AI— Cross-dataset querying with DTA, Manumissions, and Certificates of Freedom



Can you trace the journey of individuals from being listed in Domestic Traffic Ads to receiving a Certificate of Freedom, including any manumission records?



Yes, based on the available data, we can trace the journey of an individual named John Howard from being listed in Domestic Traffic Ads to receiving his Certificate of Freedom.

- Domestic Traffic Ads:** John Howard was listed for sale in a Domestic Traffic Ad on **August 15, 1829**. You can view the **original ad** through the following **source: 1829 Domestic Traffic Ad for John Howard**.
- Manumission Records:** Approximately **a year later**, his Manumission Record shows that he was granted freedom on **September 28, 1830**.
- Certificate of Freedom:** Subsequently, he was officially recognized as a free individual through his **Certificate of Freedom dated October 25, 1830**.

This progression strongly suggests that John Howard transitioned from being enslaved, being listed for sale, gaining his freedom, and then receiving official recognition of that freedom. This kind of detailed tracing can further our understanding of the process of liberation during this period.

Generative AI explanation behind the response

What the AI did:

The AI interpreted your question: 'Can you trace the journey of individuals from being listed in Domestic Traffic Ads to receiving a Certificate of Freedom, including any manumission records?', generated a Cypher query to explore graph relationships in the Knowledge Graph, executed it against Neo4j, and composed an answer from the query results.

This progression strongly suggests that John Howard transitioned from being enslaved, being listed for sale, gaining his freedom, and then receiving official recognition of that freedom. This kind of detailed tracing can further our understanding of the process of liberation during this period.

Generative AI explanation behind the response

What the AI did:

The AI interpreted your question: 'Can you trace the journey of individuals from being listed in Domestic Traffic Ads to receiving a Certificate of Freedom, including any manumission records?', generated a Cypher query to explore graph relationships in the Knowledge Graph, executed it against Neo4j, and composed an answer from the query results.

Internal Logs (Advanced):

> Entering new GraphCypherQAChain chain...

Generated Cypher:

```
cypher
MATCH (e:EnslavedPerson)-[:LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN]->(ad:DomesticTrafficAd)
(e)-[:MANUMITTED_THROUGH]->(m:ManumissionRecord),
(e)-[:WAS_GRANTED_FREEDOM_THROUGH]->(f:CertificateOfFreedom)
RETURN e.first_name AS FirstName,
e.last_name AS LastName,
ad.Ad_Image_Metadata AS AdImageLink,
ad.Sale_Date AS SaleDate,
m.DateManumitted AS ManumissionDate,
f.date_freed AS FreedomDate
ORDER BY FirstName, LastName, SaleDate
```

Full Context:

```
[{'FirstName': 'John', 'LastName': 'Howard', 'AdImageLink': 'http://msa
```

> Finished chain.

```
DEBUG: Chain Result: {'query': 'Can you trace the journey of individual
```

Clear Chat

Figure 40, 41: ChatLoS v3's KG RAG-based answer for life tracing questions



Sheriff's Sale.
Will be sold on Saturday the 15th
day of August next, at the Court house
Door in Centreville,
Between the hours of 10 & 5 o'clock
of that day, for Cash
One Negroman, named
JOHN HOWARD.
(For a Term of yrs, the property of
Stephen Boon, taken in Execution, by
Virtue of a Writ Fi. Fa. to me di-
rected, at the S of Wm. Reed, a-
gainst said Boon.
Thos. Aschom.
July 25th Sheriff.

Figure 42: ChatLoS v3's answer source linking to the traceable ad link on MSA's website

Future Work - Advancing KG-RAG with Agentic AI and Community Integration

- Expand KG schema:
 - include skills, occupations, aliases to resolve missing field contradictions
- Implement GPT Actions:
 - enable real-time interactions with MSA databases for live updates
- Use AMCP (Agentic Model Context Protocol):
 - manage multi-step archival workflows across datasets
- Introduce real-time user feedback loop:
 - thumbs up/down, comments, source flagging
- Flipped prompting for clarification:
 - AI prompts users when ambiguity is detected
- Phased rollout with community-led testing: archivists, genealogists, and students



Conclusion

**Thanks
very
much!**

