Teaching Computational Archival Science (CAS):

Context, Pedagogy, and Future Directions in Metadata

Richard **MARCIANO**
University of Maryland
marciano@umd.edu

Advanced **I**nformation **C**ollaboratory
https://ai-collaboratory.net/cas/

October 21, 2024

Dublin Core™ Metadata Initiative (DCMI) 2024:
Metadata Innovation: Trust, Transformation, and Humanity

Advanced Information Collaboratory

UNIVERSITY OF MARYLAND

DCMI 2024

This talk builds on an international collaboration: US/Canada

Professor
MAS Program Chair
Blockchain@UBC Cluster Lead

Teaching Computational Archival Science:

Context, Pedagogy & Future Directions

Victoria L. Lemieux (U. British Columbia, CANADA -- School of Information)
Richard Marciano (U. Maryland, USA -- College of Information)

Submitted to ***iConference 2025***

https://ai-collaboratory.net/wp-content/uploads/2024/09/Teaching-Computational-Archival-Science-Context-Pedagogy-and-Future-Directions-2024-submitted.pdf

Where: We illustrate the introduction of CAS into graduate archival training: *case studies*

a. University of Maryland
b. University of British Columbia
c. Building and Sustaining Educator Networks

We propose CAS graduate competencies

Computational Training Initiatives:

U.S.:

- I. TALENT Network (2018-2026):
  - 10 INFO Schools
- II. LEADING Network (2018-2025):
  - Trained 89 Fellows
  - Relied on 26 Mentors
- III. DCIP Certificate Program (2019-2024):
  - 60+ Students
- IV. DCIC Center (2015-2020):
  - 300 Students

INTERNATIONAL:

- V. Advanced Information Collaboratory - AIC (2020-present):
  - 50+ Partners

Rest of Talk:
Metadata to support GenAI & LLMs

- **OAIS & Computational Archival Processing**
- **Computational Archival Science** (CAS)
- **GraphRAG** (Retrieval Augmented Generation with Graphs)

I. TALENT Network

Training of Archival & Library Educators w. iNnovative Technologies

"Modernizing **MLIS** Teaching"

Spelman College Fall DATATHON (Nov. 1, 2024)

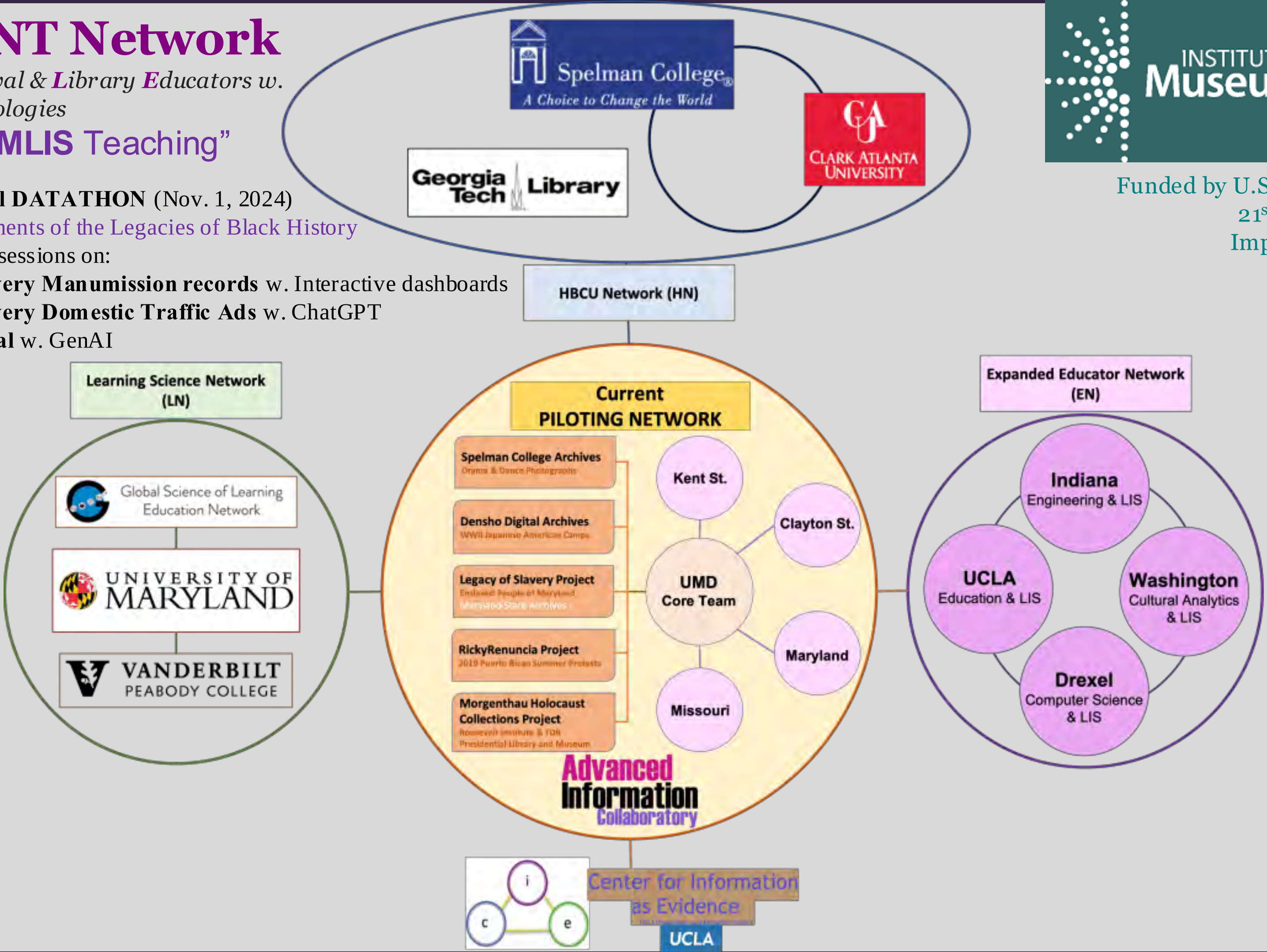
Computational Treatments of the Legacies of Black History

Featuring 3 one-hour sessions on:

- **Legacy of Slavery Manumission records** w. Interactive dashboards
- **Legacy of Slavery Domestic Traffic Ads** w. ChatGPT
- **Urban Renewal** w. GenAI

INSTITUTE of Museum and Library SERVICES

Funded by U.S. IMLS Laura Bush 21st Century Librarian Implementation Grant


Spelman College
A Choice to Change the World
Clark Atlanta University
Georgia Tech Library
HBCU Network (HN)
Learning Science Network (LN)
Global Science of Learning Education Network
University of Maryland
Vanderbilt Peabody College
Current PILOTING NETWORK
Spelman College Archives
Drama & Dance Photographs
Densho Digital Archives
WWII Japanese American Camps
Legacy of Slavery Project
Enslaved People of Maryland
Maryland State Archives
RickyRenuncia Project
2019 Puerto Rican Summer Protests
Morgenthau Holocaust Collections Project
Roosevelt Institute & FDR Presidential Library and Museum
UMD Core Team
Kent St.
Clayton St.
Maryland
Missouri
Advanced Information Collaboratory
Expanded Educator Network (EN)
Indiana
Engineering & LIS
UCLA
Education & LIS
Washington
Cultural Analytics & LIS
Drexel
Computer Science & LIS
i
c
e
Center for Information as Evidence
UCLA

INSTITUTE of Museum and Library SERVICES

1. IMLS 2018-2020: CT-LASER

Developing a Computational Framework for Library and Archival Education ($100K)

Planning grant.

https://ai-collaboratory.net/wp-content/uploads/2020/11/Final_Report_r.pdf

2. IMLS 2020-2024: PILOTING Network

Pilot Study with 4 US iSchools & 5 Archives ($300K)

Piloting grant.

https://ai-collaboratory.net/projects/piloting-network/

3. IMLS 2022-2025: TALENT Network

Promoting the Training of Archival & Library Educators w. iNnovative Technologies ($400K)

National implementation grant.

https://ai-collaboratory.net/projects/talent-network/

4. IMLS 2024-2026: GenAI-4-Arch

Harnessing **Generative AI** to Support Exploration and Discovery in Library and **Archival Collections** ($194K)

https://www.imls.gov/grants/awarded/lg-256565-ols-24

HERITAGE-AI Initiative:

Harnessing **E**nhanced **R**esearch & **I**nstructional **T**echnologies for **A**rchival **G**enerative **E**xploration - using **AI**

https://heritage-ai.org/

II. LEADING: LIS Education And Data Science Integrated Network Group *[IMLS Drexel: 2020-2025]* **($887K)**

LEADS: Library Education And Data Science *[IMLS Drexel: 2017-2020]* **($313K)**

- 14 core team members Drexel iSchool faculty and partners leads UCSD, OCLC, and Univ. of New Mexico
- 26 mentors from leading libraries, archives, data/research centers and agencies across the U.S.

https://mrc.cci.drexel.edu/leading/

"Enhancing **Doctoral** and **Early Career** Professional Training in Data Science"

→ 89 Fellows

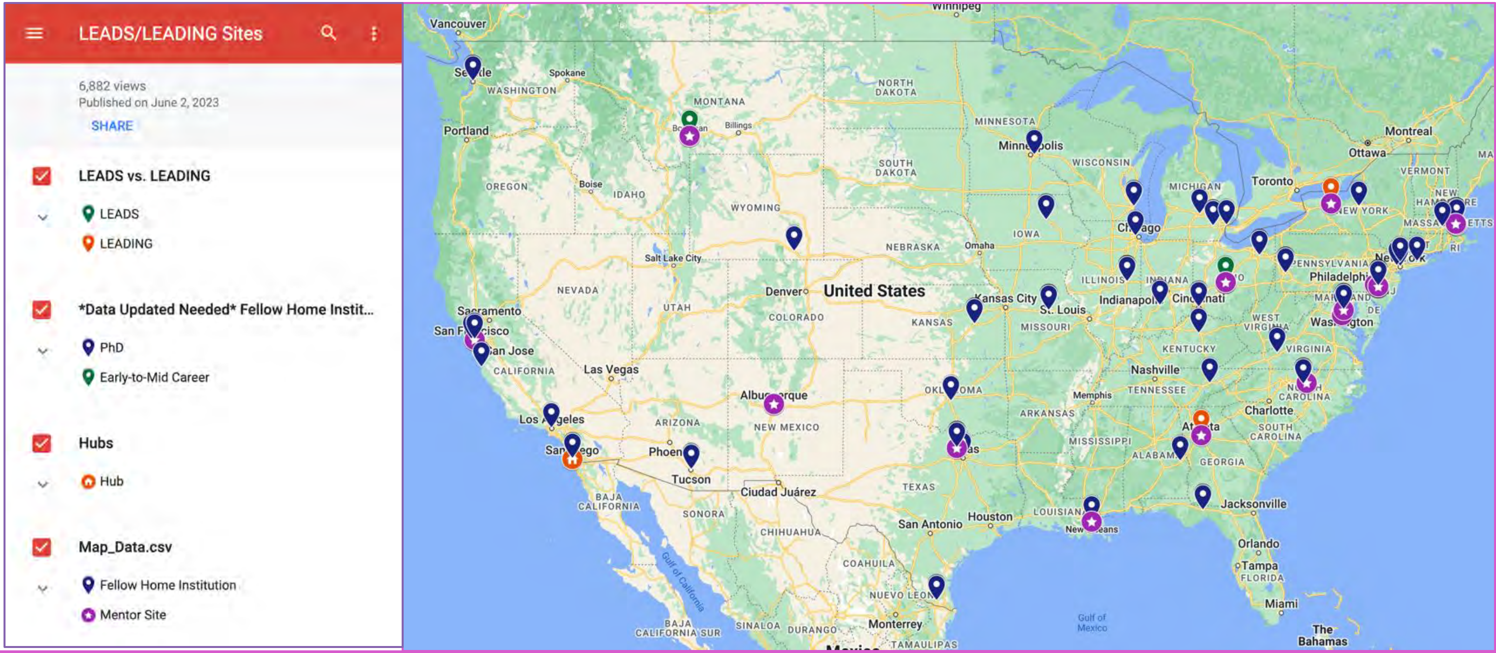
INSTITUTE of Museum and Library SERVICES

Jane Greenberg
Professor @ Drexel U.

Director @ Metadata Research Center

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

Co-Chair @ DCMI 2024
Invited Panels & Talks



LEADS/LEADING Sites

6,882 views
Published on June 2, 2023

SHARE

☑ LEADS vs. LEADING
- LEADS
- LEADING

☑ *Data Updated Needed* Fellow Home Instit...
- PhD
- Early-to-Mid Career

☑ Hubs
- Hub

☑ Map_Data.csv
- Fellow Home Institution
- Mentor Site

CURRENT / RECENT PROJECTS

EDUCATION: $2.2M

- Developing a Computational Framework for Library and Archival Education *[IMLS: 2018-2020]*
- Piloting a Collaborative Network for Integrating CT into Library and Archival Ed. and Practice *[IMLS: 2020-2024]*
- Training of Archival & Library Educators w. iNnovative Technologies *[IMLS: 2022-2023]*
- LIS Education And Data Science Integrated Network Group, LEADING *[IMLS Drexel: 2020-2025]* **($887K)**
- Library Education And Data Science, LEADS *[IMLS Drexel: 2017-2020]* **($313K)**

INFRASTRUCTURE: $14M

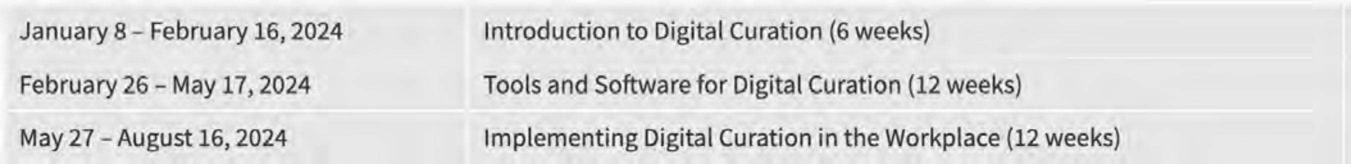
- WIN: a Window Into Neuroregulation *[NSF Convergence: 2019-2024]*
- Developing a Digital Asset Management System for Additive Manufacturing *[ARL: 2020-2024]*
- Developing a Digital Asset Management System for the Mary McLeod Bethune Historic Site *[NPS: 2019-2022]*
- Improving Fedora 4 to Work with Web-Scale Storage and Services, DRASTIC *[IMLS: 2017-2020]*
- Brown Dog: "Making Sense of Billion-Record Archives" with the NCSA) *[NSF: 2013-2018]*

SOCIAL JUSTICE, HUMAN RIGHTS, CULTURAL HERITAGE: $3M

- Using AI and ML to Optimize Information Discovery in Under-utilized, Holocaust-related Records *[Kurtz Foundation]*
- Harnessing Generative AI to Support Exploration and Discovery in Library and Archival Collections *[IMLS proposal]*
- International Research Portal for Holocaust-Era Cultural Property *[Kurtz Foundation]*
- Measuring the Impact of Urban Renewal *[NSF]*
- Computational Thinking to Unlock the Japanese American WWII Camp Experience *[UMD-FIA]*
- Computational Treatments to re-member the Legacy of Slavery (CT-LOS) *[Kurtz Foundation]*
- Testbed for the Redlining Archives of California's Exclusionary Spaces (T-RACES) *[IMLS]*
- Mapping Inequality – Redlining in New Deal America *[U. Richmond Mellon]*

III. DCIP: Digital Curation for **Information Professionals**

Certificate Program



January 8 – February 16, 2024	Introduction to Digital Curation (6 weeks)
February 26 – May 17, 2024	Tools and Software for Digital Curation (12 weeks)
May 27 – August 16, 2024	Implementing Digital Curation in the Workplace (12 weeks)

INSTITUTE of Museum and Library SERVICES

Over 60 projects

- https://ischool.umd.edu/academics/certificates-non-degree-study/digital-curation-for-information-professionals-certificate/

Examples of Student Projects:

- 2024: https://ai-collaboratory.net/2024/09/03/2024-dcip-cohort-presents-their-capstone-projects/
- 2023: https://ai-collaboratory.net/2023/08/14/2023-dcip-cohort-presents-their-capstone-projects/
  - *Synthetic data and generative AI: an interactive learning experience*
  - *Digital preservation of legacy file formats*
  - *Buried While Black – Payne Cemetery: The disinterment of a historically black cemetery in Washington, DC*
  - *Exploring Indexing Methods for Handwritten Text*
  - *Columbia MD Archives digital curation manual*

- 2022: https://ai-collaboratory.net/2022/09/07/2022-dcip-cohort-presents-their-capstone-projects/
- 2021: https://ai-collaboratory.net/2021/08/23/dcip-cohort-2021-presents-capstone-projects/

IV. DCIC: 2015-2020

CAS

Computational Archival Science

The DCIC is pioneering advances in *computational treatments of archival and cultural content*.

See our CAS portal for the latest developments:
http://dcicblog.umd.edu/cas/

What is CAS?

An interdisciplinary field concerned with the application of computational methods and resources to large-scale rec-ords /archives processing, analysis, stor-age, long-term preservation, and ac-cess, with the aim of improving efficien-cy, productivity and precision in support of appraisal, arrangement and descrip-tion, preservation and access decisions, and engaging and undertaking re-search with archival materials.

CAS Founding Partners:

Richard Marciano, U. Maryland
Mark Hedges, King's College London (UK)
Vicki Lemieux, U. British Columbia (Canada)
Maria Esteva, Texas Advanced Computing Center (TACC)
Michael Kurtz, U. Maryland
Bill Underwood, U. Maryland
Greg Jansen, U. Maryland
Mark Conrad, National Archives and Records Administration (NARA)

curatelab

Hornbake South 4110

Digital lab for group learning, collaborative design, and hands-on digital curation project development (23 seats, 3 interactive screens, 12 workstations with 12TB of storage).

digitizationlab

Hornbake South 4110D

Document scanning, image manipulation, and archival ingestion facility for group projects.

serverfarm

Atlantic Building

On-campus virtual machine farm for research data processing, storage, and hosting (15TB storage, 2 Dell servers, VMWare-powered).

cloudlab

Amazon Cloud

Dashboard-enabled virtual computing lab in the cloud for creating Windows/Ubuntu instances using Amazon Web Services (AWS).

dataCave

UMD Cyberinfrastructure Center at the Rivertech Bldg

DRAS-TIC

Digital **R**epository **A**t **S**cale **T**hat **I**nvites **C**omputa-tion (**T**o **I**mprove **C**ollections): a petascale archival storage and preservation repository (based on the **DRAS-TIC** open-source software [NoSQL Cassandra database] and computational infrastructure (4 Dell nodes).

dcic

digital curation innovation center

http://dcic.umd.edu

Mission:

Be a leader in the digital curation research and educational fields, and foster interdisciplinary collaborations using Big Records and Archival Analytics with public / industry / government partnerships.

Goals:

Sponsor interdisciplinary projects that explore the integration of archival research data, user-contributed data, and technology to generate new forms of analysis and historical research engagement, particularly in the arenas of social justice, human rights, and cultural heritage.

Motto:

"Integrating Education and Research"

UNIVERSITY OF MARYLAND 18 56
COLLEGE OF
INFORMATION
STUDIES

David Weintrop:

- CT-STEM Practices Taxonomy

CITATION

Bill Underwood:

- CAS#4:

Analysis of the remaining eleven workshop papers indicates that the research that they report also involves CT.

https://ai-collaboratory.net/wp-content/uploads/2020/02/16_OpenMic_Bill-Underwood.pdf

- CT-LASER Practice:

Motivation for Integrating CT into UMD MLIS program in Library and Archival Studies, with examples of CT Practices being used in Archival Studies Research

https://ai-collaboratory.net/wp-content/uploads/2020/04/Underwood_CompThinkInArchResearch.pdf

Computational archival science is a blend of: (1) computational & (2) archival thinking.

DP
Data Practices

- Collecting Data
- Creating Data
- Manipulating Data
- Analyzing Data
- Visualizing Data

MS
Modeling & Simulation Practices

- Using Computational Models to Understand a Concept
- Using Computational Models to Find and Test Solutions
- Assessing Computational Models
- Designing Computational Models
- Constructing Computational Models

CPS
Computational Problem Solving Practices

- Preparing Problems for Computational Solutions
- Programming
- Choosing Effective Computational Tools
- Assessing Different Approaches/Solutions to a Problem
- Developing Modular Computational Solutions
- Creating Computational Abstractions
- Troubleshooting and Debugging

ST
Systems Thinking Practices

- Investigating a Complex System as a Whole
- Understanding the Relationships within a System
- Thinking in Levels
- Communicating Information about a System
- Defining Systems and Managing Complexity

Example 1: Fall 2019

Experiential, Interdisciplinary & Team-based Learning:

Computational Thinking in Archives

"Reframing Digital Curation Practices through a Computational Thinking Framework"

Richard Marciano, et al., 2019 IEEE International Conference on Big Data, 4th CAS Workshop, Dec. 11, 2019, Los Angeles, CA.

https://ai-collaboratory.net/wp-content/uploads/2020/04/ReframingDC-UsingCT_final.pdf

[Runaway Slave Ads]

[Certificates of Freedom]

[Manumissions]

[Cemetery Records]

[LEADS Fellows]

October 28-29, 2019: "Datathon" at the Maryland State Archives

https://ai-collaboratory.net/projects/legacy-of-slavery/student-led-datathon-at-the-maryland-state-archives/

- **Paper**: Gnanasekaran, R.K. and Marciano, R., (2021). *Piloting Data Science Learning Platforms through the Development of Cloud-based interactive Digital Computational Notebooks.* https://ai-collaboratory.net/wp-content/uploads/2021/10/ISGC2021_Gnanasekaran_Marciano.pdf.
- **Video**: https://www.youtube.com/watch?v=cNBc0AY-r-k
- **Jupyter Notebook**: https://cases.umd.edu/github/cases-umd/Legacy-of-Slavery/blob/master/index.ipynb

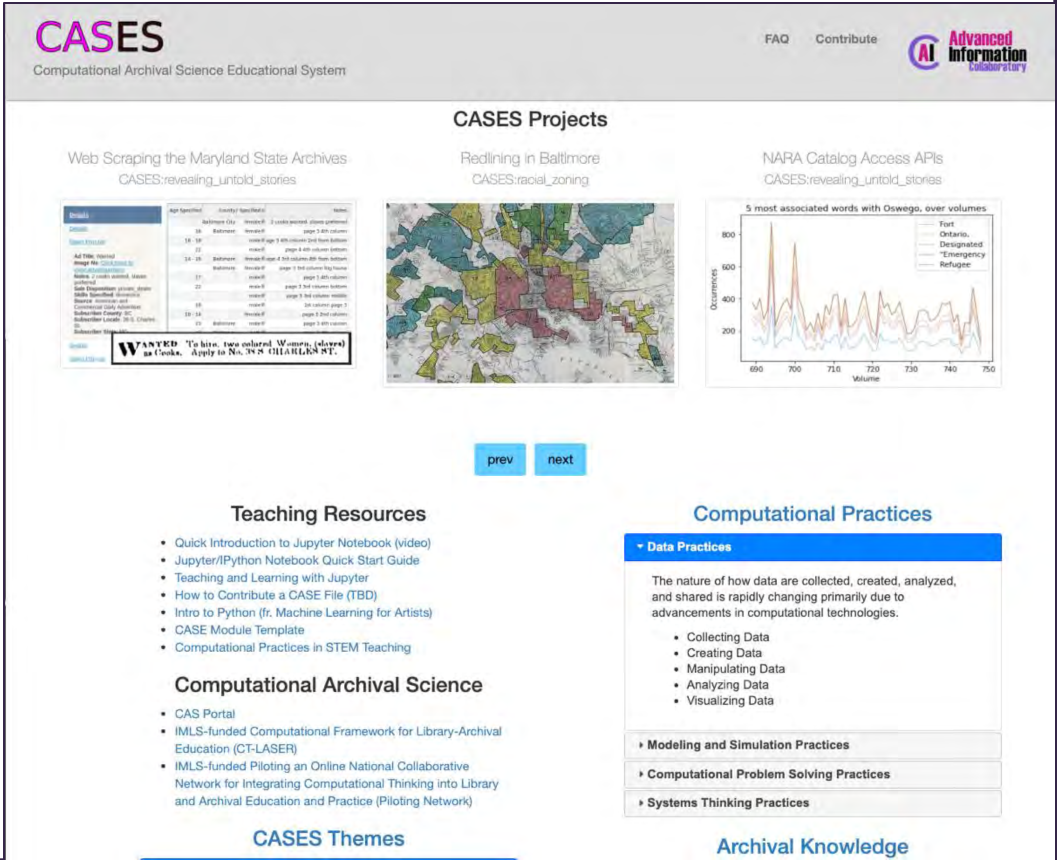
(Jupyter) Digital Notebooks

Teaching and Learning with Jupyter
May 8, 2019
https://jupyter4edu.github.io/jupyter-edu-book/

Educators are rapidly adopting Jupyter Notebooks for:

* teaching
* use in the classroom
* developing teaching materials
* creating computational stories

See: https://cases.umd.edu


FAQ
Contribute
AI Advanced Information Collaboratory
WANTED To hire, two colored Women, (slaves) as Cooks. Apply to No. 38 S CHARLES ST.
5 most associated words with Oswego, over volumes
Fort
Ontario.
Designated
"Emergency
Refugee
Occurrences
800
600
400
200
690
700
710
720
730
740
750
Volume
prev
next

CASES
Computational Archival Science Educational System

CASES Projects

Web Scraping the Maryland State Archives
CASES:revealing_untold_stories

Redlining in Baltimore
CASES:racial_zoning

NARA Catalog Access APIs
CASES:revealing_untold_stories

Teaching Resources

- Quick Introduction to Jupyter Notebook (video)
- Jupyter/IPython Notebook Quick Start Guide
- Teaching and Learning with Jupyter
- How to Contribute a CASE File (TBD)
- Intro to Python (fr. Machine Learning for Artists)
- CASE Module Template
- Computational Practices in STEM Teaching

Computational Archival Science

- CAS Portal
- IMLS-funded Computational Framework for Library-Archival Education (CT-LASER)
- IMLS-funded Piloting an Online National Collaborative Network for Integrating Computational Thinking into Library and Archival Education and Practice (Piloting Network)

CASES Themes

Computational Practices

▾ Data Practices

The nature of how data are collected, created, analyzed, and shared is rapidly changing primarily due to advancements in computational technologies.

- Collecting Data
- Creating Data
- Manipulating Data
- Analyzing Data
- Visualizing Data

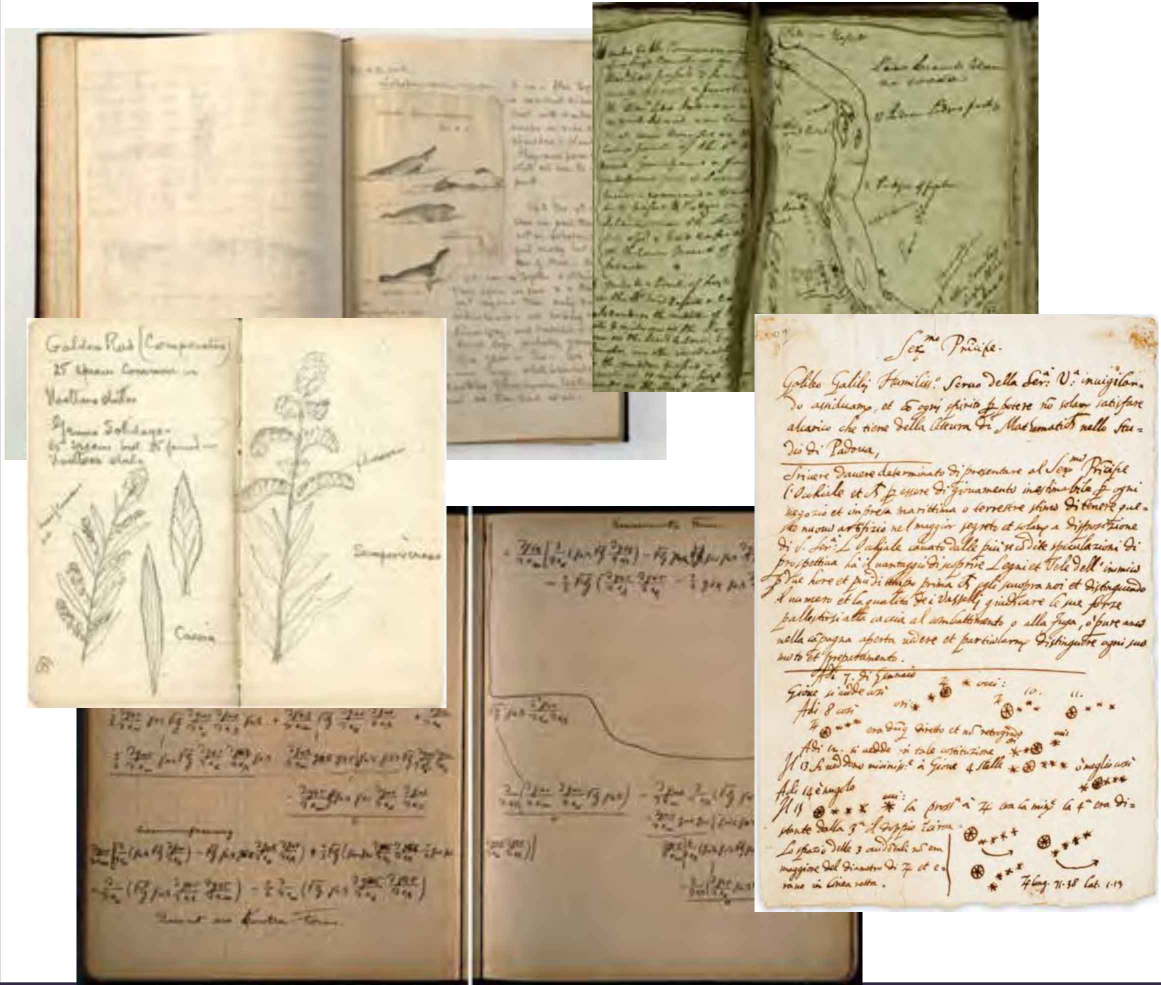
▸ Modeling and Simulation Practices

▸ Computational Problem Solving Practices

▸ Systems Thinking Practices

Archival Knowledge

Historical Lab Notebooks



Golden Rod (Compositae)
Solidago
Sempervirens
Caesia

Ser.mo Principe.

Galileo Galilei Humiliss.o Servo della Ser.tà V.a invigilando assiduam.te, et con ogni spirito p potere non solam.te satisfare al carico che tiene della Lettura di Matematiche nello Studio di Padova,

Havere determinato di presentare al Ser.mo Principe l'Occhiale et di p essere di giovamento inestimabile p ogni negozio et impresa marittima o terrestre stima di tenere questo nuovo artifizio nel maggior segreto et solam.te a disposizione di V. Ser.tà L'Occhiale cavato dalle più recondite speculazioni di prospettiva ha il vantaggio di scoprire Legni et Vele dell'inimico p due hore et più di tempo prima che egli scuopra noi et distinguendo il numero et la qualità dei Vasselli giudicare le sue forze pallestirsi alla caccia al combattimento o alla fuga, o pure ancora nella campagna aperta vedere et particolarm.te distinguere ogni suo moto et preparamento.

Adi 7. di Gennaio Giove si vedeva così ✶ ⊛ ✶ così:

Adi 8 così ⊛ ✶✶✶ era dunq. diretto et no retrogrado

10. ⊛ ✶✶ 11. ⊛ ✶✶✶

Adi 12. si vedde in tale costituzione ✶ ✶ ⊛ ✶ così

Il 13 si veddono vicinissime à Giove 4 stelle ✶ ⊛ ✶✶✶ ò meglio così ✶ ⊛ ✶✶✶

Adi 14 è nugolo

Il 15 ⊛ ✶ ✶ ✶ ✶ così: la pross.a à ♃ era la min.e la 4a era distante dalla 3a il doppio incirca.

Lo spazio delle 3 occidentali no era maggiore del diametro di ♃ et erano in linea retta.

♃ long. 71.38 lat. 1.13

Paper-based Lab Notebooks:

- Used in science research
- Represent a record of:
  - observations
  - experiments
  - ideas
  - notes
  - formulas
  - data

Electronic Lab Notebooks:

- patient medical records

1. Developing Name Registries

Creating a Names Registry

Data Practices	Modeling & Simulation Practices	Computational Problem Solving Practices	Systems Thinking Practices
Collecting Data	**Using Computational Models to Understand a Concept**	Preparing Problems for Computational Solutions	**Investigating a Complex System as a Whole**
Creating Data	**Using Computational Models to Find and Test Solutions**	**Programming**	**Understanding the Relationships within a System**
Manipulating Data	**Assessing Computational Models**	Choosing Effective Computational Tools	**Thinking in Levels**
Analyzing Data	**Designing Computational Models**	Assessing Different Approaches/Solutions to a Problem	**Communicating Information about a System**
Visualizing Data	**Constructing Computational Models**	Developing Modular Computational Solutions	Defining Systems and Managing Complexity
		Creating Computational Abstractions	
		Troubleshooting and Debugging	

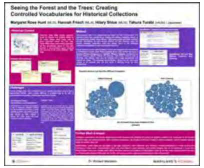
2. Integrating Vital Records

Consolidating and Analyzing Vital Records

Contextualizing Vital Records

Data Practices	Modeling & Simulation Practices	Computational Problem Solving Practices	Systems Thinking Practices
Collecting Data	Using Computational Models to Understand a Concept	**Preparing Problems for Computational Solutions**	**Investigating a Complex System as a Whole**
Creating Data	Using Computational Models to Find and Test Solutions	Programming	**Understanding the Relationships within a System**
Manipulating Data	Assessing Computational Models	Choosing Effective Computational Tools	**Thinking in Levels**
Analyzing Data	Designing Computational Models	Assessing Different Approaches/Solutions to a Problem	**Communicating Information about a System**
Visualizing Data	Constructing Computational Models	Developing Modular Computational Solutions	Defining Systems and Managing Complexity
		Creating Computational Abstractions	
		Troubleshooting and Debugging	

3. Designing Controlled Vocabularies


Seeing the Forest and the Trees: Creating Controlled Vocabularies for Historical Collections

Data Practices	Modeling & Simulation Practices	Computational Problem Solving Practices	Systems Thinking Practices
Collecting Data	**Using Computational Models to Understand a Concept**	**Preparing Problems for Computational Solutions**	**Investigating a Complex System as a Whole**
Creating Data	**Using Computational Models to Find and Test Solutions**	**Programming**	**Understanding the Relationships within a System**
Manipulating Data	Assessing Computational Models	Choosing Effective Computational Tools	**Thinking in Levels**
Analyzing Data	Designing Computational Models	Assessing Different Approaches/Solutions to a Problem	**Communicating Information about a System**
Visualizing Data	Constructing Computational Models	Developing Modular Computational Solutions	Defining Systems and Managing Complexity
		Creating Computational Abstractions	
		Troubleshooting and Debugging	

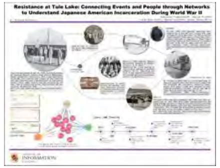
4. Mapping Events and People

Data Practices	Modeling & Simulation Practices	Computational Problem Solving Practices	Systems Thinking Practices
Collecting Data	**Using Computational Models to Understand a Concept**	**Preparing Problems for Computational Solutions**	**Investigating a Complex System as a Whole**
Creating Data	**Using Computational Models to Find and Test Solutions**	Programming	**Understanding the Relationships within a System**
Manipulating Data	Assessing Computational Models	Choosing Effective Computational Tools	**Thinking in Levels**
Analyzing Data	Designing Computational Models	Assessing Different Approaches/Solutions to a Problem	**Communicating Information about a System**
Visualizing Data	Constructing Computational Models	Developing Modular Computational Solutions	Defining Systems and Managing Complexity
		Creating Computational Abstractions	
		Troubleshooting and Debugging	

Mapped Narratives of Resistance in Tule Lake

5. Connecting Events and People

Data Practices	Modeling & Simulation Practices	Computational Problem Solving Practices	Systems Thinking Practices
Collecting Data	**Using Computational Models to Understand a Concept**	**Preparing Problems for Computational Solutions**	**Investigating a Complex System as a Whole**
Creating Data	**Using Computational Models to Find and Test Solutions**	Programming	**Understanding the Relationships within a System**
Manipulating Data	Assessing Computational Models	Choosing Effective Computational Tools	**Thinking in Levels**
Analyzing Data	Designing Computational Models	Assessing Different Approaches/Solutions to a Problem	**Communicating Information about a System**
Visualizing Data	Constructing Computational Models	Developing Modular Computational Solutions	Defining Systems and Managing Complexity
		Creating Computational Abstractions	
		Troubleshooting and Debugging	


Resistance at Tule Lake: Connecting Events and People through Networks to Understand Japanese American Incarceration During World War II

DEVELOPING NAME REGISTRIES

Andy Jose SILVA / Emery PATTERSON, Mary McKINLEY
InfoSci 2 MLIS

INTEGRATING VITAL RECORDS

James SANTOS, Genevieve KOCIENDA / Kanishka JAIN
HLS MLIS MIM

DESIGNING CONTROLLED VOCABULARIES

Tahsin TURABI / Margaret Rose HUNT, Hannah FRISCH, Hilary Szu Yin SHIUE
InfoSci / Japanese 3 MLIS

MAPPING EVENTS & PEOPLE

Connor MULLANE, Britton SCHAMS, Mirielle VASSELLI / Chenxi LIU, Jade XU
3 MLIS 2 BCIM

CONNECTING EVENTS & PEOPLE

Sarah AGARRAT / Hannah KRAUSS / Danish MIR, Mehak SURI / Debadrish PRADHAN
CS / Math MLIS 2 MIM BCIM

EVENT: Resistance at Tule Lake: A Conversation with the Filmmaker and iSchool Digital Curators (and Film Viewing)

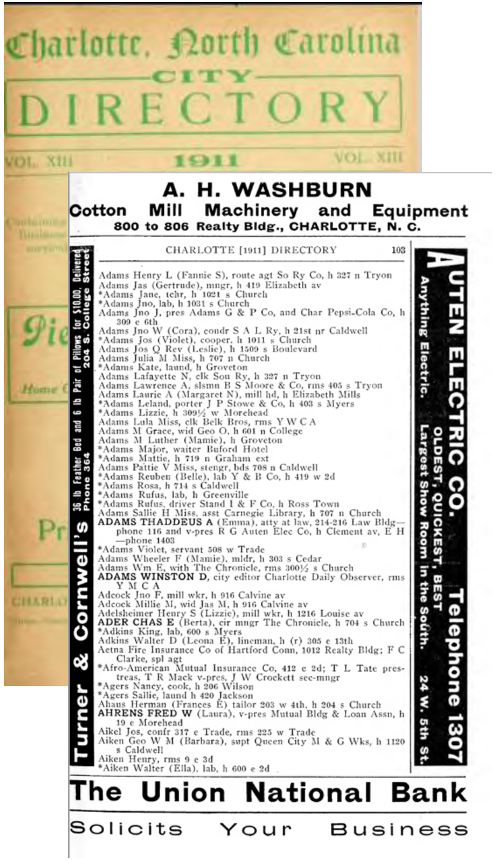
PROJECT:

Computational Thinking to Unlock the Japanese American WWII Camp Experience (w. Densho.org)

https://ai-collaboratory.net/projects/ct-ja_ww2_camps/

Example 2: Spring 2023 Graduate Course in the MLIS Program – Implementing Digital Curation

- ***Teaching and Learning with Archival Materials through the Development of Interactive Computational Notebooks",***
  P. Piety, M. Conrad, R. Marciano, I. Cornfield, E. Dallimore, R. Fettig, E. Hansen, H. Kemp, T. Turabi (2023).
  Chapter Submission for the 2023 Archives and Primary Source Handbook, peer-reviewed open-access NewPrairiePress textbook.
  Link: https://ai-collaboratory.net/wp-content/uploads/2023/10/Piety_Conrad_Marciano_et_al-FINAL.pdf



Charlotte, North Carolina
CITY
DIRECTORY
VOL. XIII 1911 VOL. XIII

A. H. WASHBURN
Cotton Mill Machinery and Equipment
800 to 806 Realty Bldg., CHARLOTTE, N. C.

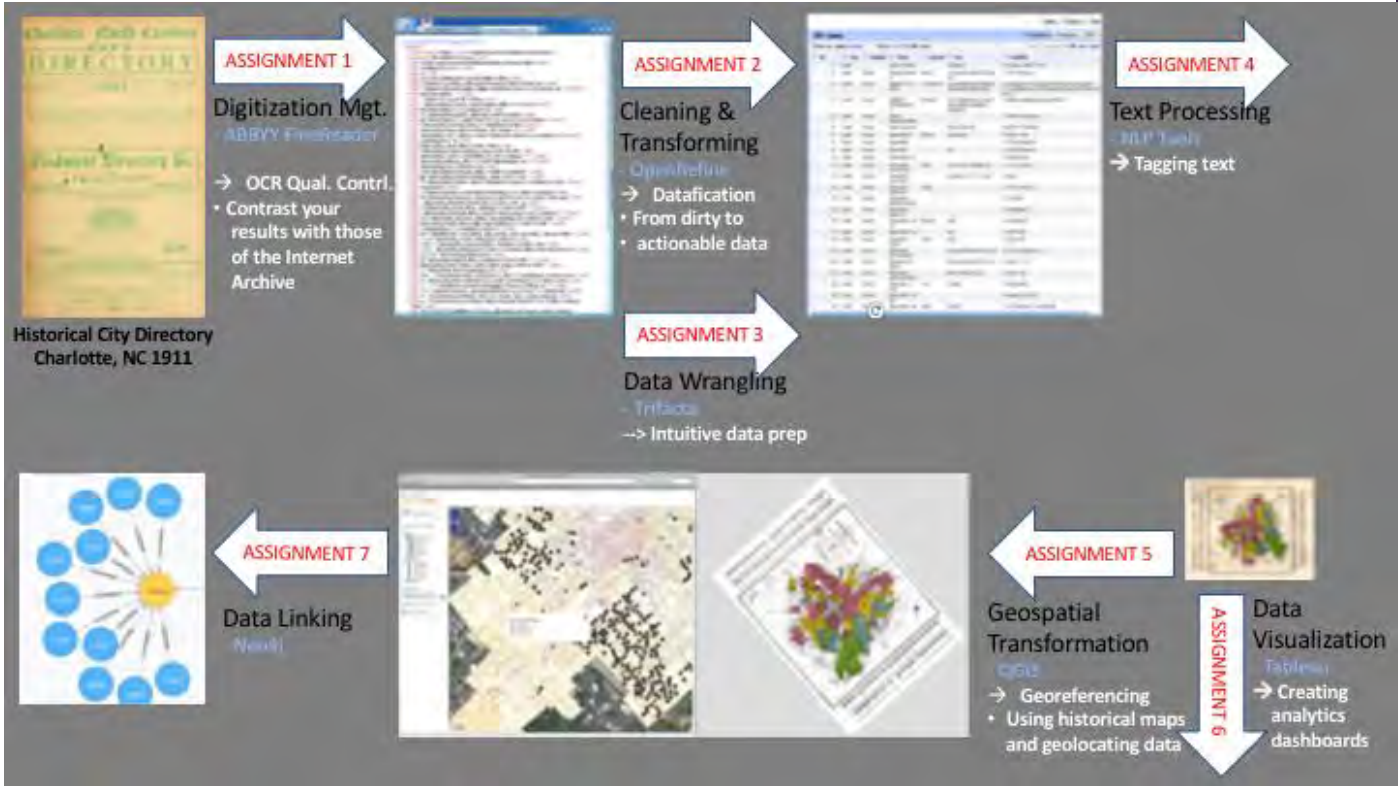
CHARLOTTE [1911] DIRECTORY 103

Adams Henry L (Fannie S), route agt So Ry Co, h 327 n Tryon
Adams Jas (Gertrude), mngr, h 419 Elizabeth av
*Adams Jane, tchr, h 1021 s Church
*Adams Jno, lab, h 1031 s Church
Adams Jno J, pres Adams G & P Co, and Char Pepsi-Cola Co, h 309 e 6th
Adams Jno W (Cora), condr S A L Ry, h 21st nr Caldwell
*Adams Jos (Violet), cooper, h 1011 s Church
Adams Jos Q Rev (Leslie), h 1509 s Boulevard
Adams Julia M Miss, h 707 n Church
*Adams Kate, laund, h Groveton
Adams Lafayette N, clk Sou Ry, h 327 n Tryon
Adams Lawrence A, slsmn B S Moore & Co, rms 405 s Tryon
Adams Laurie A (Margaret N), mill hd, h Elizabeth Mills
*Adams Leland, porter J P Stowe & Co, h 403 s Myers
*Adams Lizzie, h 309½ w Morehead
Adams Lula Miss, clk Belk Bros, rms Y W C A
Adams M Grace, wid Geo O, h 601 n College
Adams M Luther (Mamie), h Groveton
*Adams Major, waiter Buford Hotel
*Adams Mattie, h 719 n Graham ext
Adams Pattie V Miss, stengr, bds 708 n Caldwell
*Adams Reuben (Belle), lab Y & B Co, h 419 w 2d
*Adams Rosa, h 714 s Caldwell
*Adams Rufus, lab, h Greenville
*Adams Rufus, driver Stand I & F Co, h Ross Town
Adams Sallie H Miss, asst Carnegie Library, h 707 n Church
ADAMS THADDEUS A (Emma), atty at law, 214-216 Law Bldg—phone 116 and v-pres R G Auten Elec Co, h Clement av, E H—phone 1403
*Adams Violet, servant 508 w Trade
Adams Wheeler F (Mamie), mldr, h 303 s Cedar
Adams Wm E, with The Chronicle, rms 300½ s Church
ADAMS WINSTON D, city editor Charlotte Daily Observer, rms Y M C A
Adcock Jno F, mill wkr, h 916 Calvine av
Adcock Millie M, wid Jas M, h 916 Calvine av
Adelsheimer Henry S (Lizzie), mill wkr, h 1216 Louise av
ADER CHAS E (Berta), cir mngr The Chronicle, h 704 s Church
*Adkins King, lab, 600 s Myers
Adkins Walter D (Leona E), lineman, h (r) 305 e 13th
Aetna Fire Insurance Co of Hartford Conn, 1012 Realty Bldg; F C Clarke, spl agt
*Afro-American Mutual Insurance Co, 412 e 2d; T L Tate pres-treas, T R Mack v-pres, J W Crockett sec-mngr
*Agers Nancy, cook, h 206 Wilson
*Agers Sallie, laund h 420 Jackson
Ahaus Herman (Frances E) tailor 203 w 4th, h 204 s Church
AHRENS FRED W (Laura), v-pres Mutual Bldg & Loan Assn, h 19 e Morehead
Aikel Jos, confr 317 e Trade, rms 225 w Trade
Aiken Geo W M (Barbara), supt Queen City M & G Wks, h 1120 s Caldwell
Aiken Henry, rms 9 e 3d
*Aiken Walter (Ella), lab, h 600 e 2d

Turner & Cornwell's 36 lb Feather Bed and 6 lb Pair of Pillows for $10.00, Delivered Phone 364 204 S. College Street

AUTEN ELECTRIC CO. Telephone 1307
Anything Electric. OLDEST, QUICKEST, BEST Largest Show Room in the South. 24 W. 5th St.

The Union National Bank
Solicits Your Business



Historical City Directory
Charlotte, NC 1911

ASSIGNMENT 1
Digitization Mgt.
- ABBYY FineReader
→ OCR Qual. Contrl.
- Contrast your results with those of the Internet Archive

ASSIGNMENT 2
Cleaning & Transforming
- OpenRefine
→ Datafication
- From dirty to
- actionable data

ASSIGNMENT 3
Data Wrangling
- Trifacta
--> Intuitive data prep

ASSIGNMENT 4
Text Processing
- NLP Tools
→ Tagging text

ASSIGNMENT 5
Geospatial Transformation
- QGIS
→ Georeferencing
- Using historical maps and geolocating data

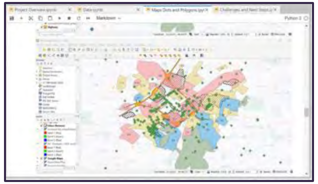
ASSIGNMENT 6
Data Visualization
- Tableau
→ Creating analytics dashboards

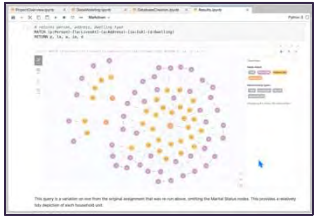
ASSIGNMENT 7
Data Linking
- Neo4j

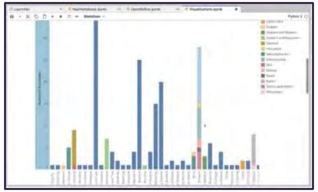
See: **videos and notebooks** at: https://ai-collaboratory.net/2023/05/11/may-11-2023-computational-storytelling-datathon/

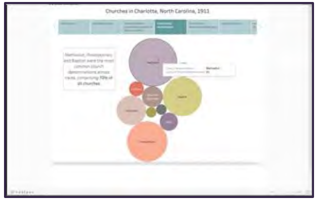
Table 2 - Comparison of student final projects

Student Project Title	Computational Thinking																						Computational Technologies									
	Data Practices					Modeling & Simulation					Computational Problem Solving							Systems Thinking Practices					Tools							Technical Sophistication		
	CLD	CRD	MD	AD	VD	UaC	F&T	ACM	DCM	CCM	PCS	PRG	CET	ADS	DMS	CCA	TAD	ISW	URS	TIL	CIS	DSC	OCR	Tableau	Excel	Neo4J	Open Refine	QGIS	Pandas	Code Boxes	Dataset Scale	Use of Code
A. Expanding the Network: Modeling Relationships with Neo4j			X						X	X							X		X						X	X				No	Small	1
B. Revisualizing Geographic Disparities: Examining Trends in Racial and Economic Inequality on the Streets w/o GIS			X	X	X	X					X			X					X	X							X			No	Full	1
C. Mad or Madam: Investigating an Undefined Data Term	X	X	X	X	X																			X			X			No	Full	1
D. Race, Marriage, and Profession: Data at Scale Test Case			X	X	X	X										X				X	X			X			X			Tableau	Full	2
E. Building a Bigger Picture: A Case Study of Combining the General City and Business Directories	X		X	X	X												X		X	X			X	X						Tableau	Full	2
F. Religious Life in 1911 Charlotte, NC		X	X	X	X												X							X			X			Tableau	Full	2
G. Gender, Race, and Archival Silences		X	X	X	X																			X			X			Tableau	Full	2
H. Conceptualizing Prosperity: A Case Study Analyzing Housing through Job Types				X	X	X												X	X	X	X			X			X			Tableau	Full	2
I. Visualizing Neighborhood Demographics			X	X	X								X						X	X				X			X	X		Tableau	Full	2
J. Mapping Over Time in Charlotte NC: Population, Redlining, and Urban Renewal		X	X		X	X			X	X	X		X				X	X	X	X	X			X			X		X	Python / Pandas Lib.	Full	3









The Future is Here & Now

- Fundamental changes to the way we acquire, manage, and present cultural collections
- COVID-19: impact on access to archives, libraries, museums

HERE:

- **US: OMB M-23-07 Update to Transition to Electronic Records**
  - **By June 30, 2024,** Federal agencies must manage all permanent records in an electronic format for eventual transfer and accessioning by NARA.
  - **After June 30, 2024**, Federal agencies must transfer all permanent records in an electronic format and with appropriate metadata.
- **NARA Budget Estimates 622 Years to Process Backlogged FOIA Requests at Just Two Presidential Libraries**
  - A "FOIA backlog of an estimated 183 million pages at the George W. Bush Library and [a] 128-million-page backlog at the Barack Obama Library" alone. At current rates, it would take NARA 622 years to declassify the pending declassification requests just these two presidential libraries.
  - NARA presently holds 13.5 billion pages, **only about 2% of which is in digitized form**. This is just one among many thousands of archival repositories; with more than 25,000 such repositories in just the United States.

NOW:

- **"An IMPERATIVE for educating the archivists and records managers of the future for the digital world"** *[Mark Conrad]*

LAUNCH OF THE AIC COLLABORATORY

https://ai-collaboratory.net

Advanced Information Collaboratory

V.

The AIC was launched at The Alan Turing Institute in London, UK on Jan. 20, 2020. It brings together partners from leading academic and cultural institutions from six continents. Its goals are to:

1. EXPLORE the opportunities and challenges of "disruptive technologies" for archives and records management (digital curation, machine learning, AI, etc.).
2. PURSUE multidisciplinary collaborations to share relevant knowledge across domains.
3. LEVERAGE the latest technologies to unlock the hidden information in massive stores of records.
4. TRAIN current and future generations of information professionals to think computationally and rapidly adapt new technologies to meet their increasingly large and complex workloads.
5. PROMOTE ethical information access and use.

The Alan Turing Institute
Computational archival science (CAS) symposium
Towards a transatlantic programme
Home

AIC Founding Partners:

Dr. Richard Marciano
Professor
UMD iSchool (US)

Mark Conrad
Archives Specialist
NARA - former (US)

Dr. Eirini Goudourali
Head of Dig. Research Progs.
TNA (UK)

Dr. Jane Greenberg
Professor
Dir. Metadata Research Center
Drexel U. (US)

Dr. Mark Hedges
Dep. of Dig. Hum.
King's College London (UK)

Greg Jansen
Senior Res. Soft. Architect
UMD iSchool (US)

Dr. Michael Kurtz
Asst. Archivist for Rec. Services
NARA - former (US)

Dr. Victoria Lemieux
Assoc. Professor
Blockchain@UBC Cluster
lead (Canada)

Dr. Bill Underwood
Res. Scientist
GTRI Res. Sci (former)
UMD iSchool (US)

Dr. Lyneise Williams
Associate Prof. Art History
Founder VERA Collaborative
UNC Chapel Hill (US)

Advanced Information Collaboratory

North America:

- **MEDiAL Lab @U. Maryland:** Dr. Phil Piety
- **NARA (former):** Bruce Ambacher
- **UCLA:** Dr. Anne Gilliland
- **Kent State U.:** Dr. Karen Gracy
- **U. Missouri:** Dr. Sarah Buchanan
- **Clayton State U.:** Dr. Joshua Kitchens
- **The Smithsonian Institutions (NMAH):** Bob Horton
- **Harvard Library:** Ceilyn Boyd
- **UC Santa Barbara:** Marisol Ramos
- **UC San Diego:** Dr. Andrea Chiba
- **US Holocaust Memorial Museum:** Michael Levy
- **Densho.org:** Geoff Froh
- **Maryland State Archives:** Chris Haley & Maya Davis
- **Spelman College:** Holly Smith
- **Puerto Rican Spring Project:** Marison Ramos, Irmarie Fraticelli, Joel Blanco

South America:

- **U. de Brasilia:** Dr. Cláudio Gottschalg-Duque

UK:

- **Loughborough U.:** Lise Jaillant
- **The Alan Turing Institute:** David Beavan
- **UK TNA:** Pip Wilcox, Mark Bell, Paul Young, Jenny Bunn & Sonia Ranade
- **Oxford U.:** Dr. David De Roure
- **European Holocaust Research Infrastructure:** Dr. Reto Speck

Europe:

- **Hamburg U. Archives:** Francesco Gelati
- **University of Amsterdam:** Dr. Tobias Blanke
- **INESC-ID Portugal:** Dr. Diogo Proença

Africa:

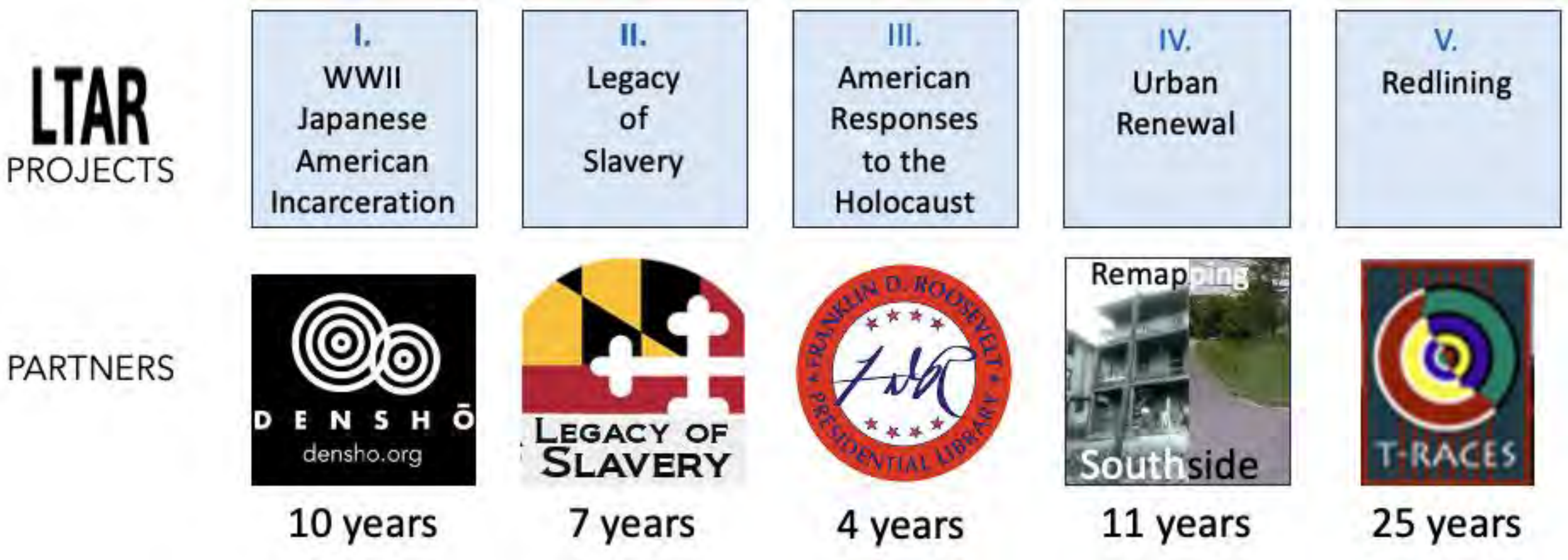
- **U. South Africa:** Dr. Shadrack Katuu

Asia:

- **Central U. of Gujarat (India):** Dr. Bhakti Gala
- **Centre for Dev. of Advanced Computation (India):** Dr. D. Katre
- **Indian Inst. of Management:** Dr. H. Anil Kumar
- **Kyushu U. (Japan):** Dr. Yoichi Tomiura & Dr. Emi Ishita

Australia:

- **U. Canberra:** Dr. Tim Sherratt



LTAR PROJECTS	I. WWII Japanese American Incarceration	II. Legacy of Slavery	III. American Responses to the Holocaust	IV. Urban Renewal	V. Redlining
PARTNERS	DENSHŌ densho.org	LEGACY OF SLAVERY	FRANKLIN D. ROOSEVELT PRESIDENTIAL LIBRARY	Remapping Southside	T-RACES
	10 years	7 years	4 years	11 years	25 years

Example:

- **SUMMER 2024: MLIS Students Engage w. Innovative Technologies to Explore the Future Processing of Archival Collections through *spatial, graph & genAI* Techniques:** https://ai-collaboratory.net/2024/07/09/summer2024/
- **Victoria Lemieux @ UBC:** "Archival Competencies Framework for Training in AI/ML"

  (using concepts of trustworthiness and authenticity of records)

How Do Computational Processes Touch Upon Archival Work?

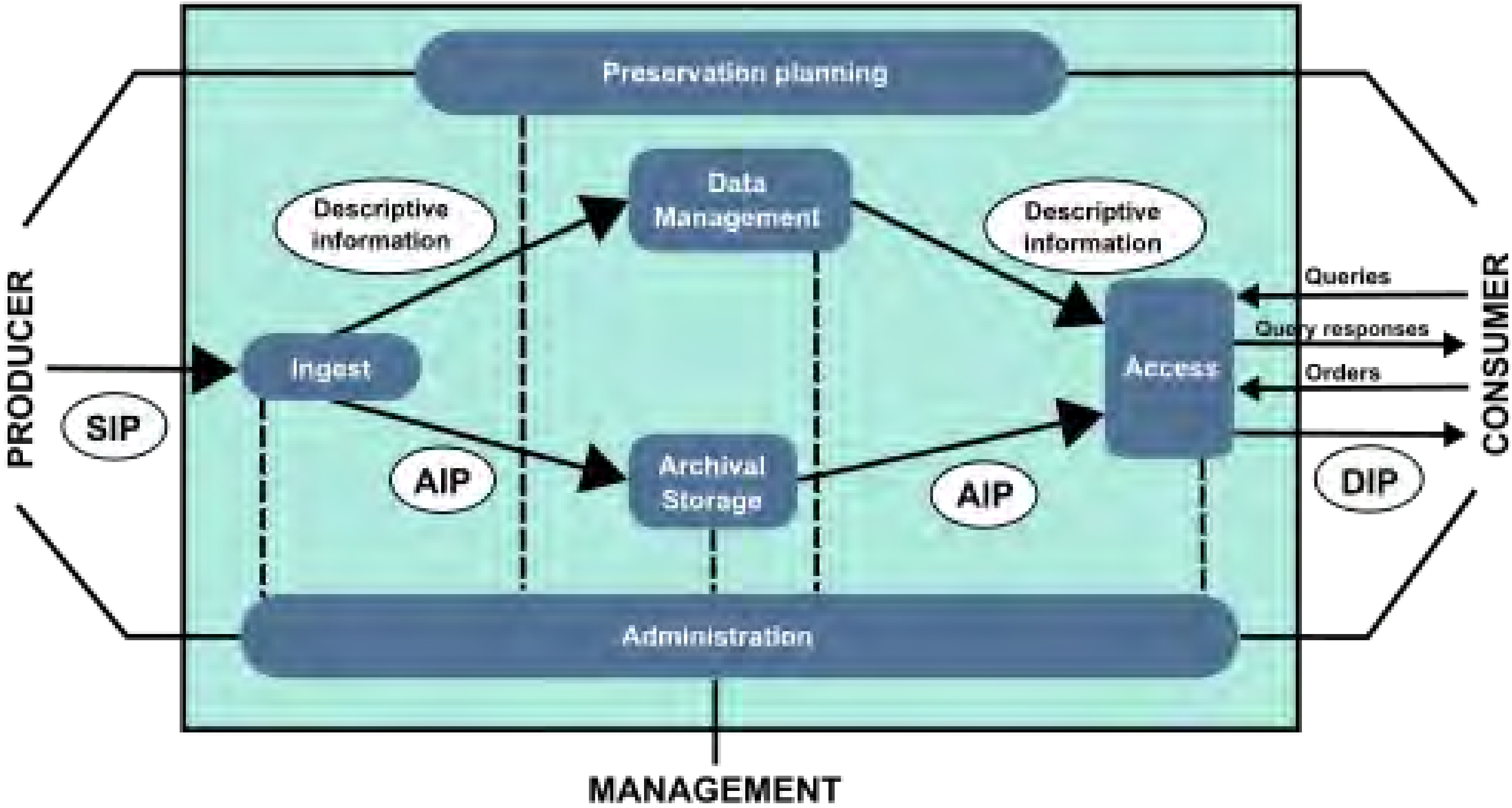
Preparing Archivists in Computational Thinking & Innovative Technologies
https://www.youtube.com/watch?v=ZdLJHQLbR4k

Anne Gilliland (UCLA) & Richard Marciano (UMD),

International Conference of **“Technology, Society, Humanities: Digital Intelligence Empowers the Modernization of Archival Work”**

Shandong U., China. Oct. 26, 2023.

Open Archival Information System (OAIS)



Mathieualexhache (original work); Mess (SVG conversion & English translation)
CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

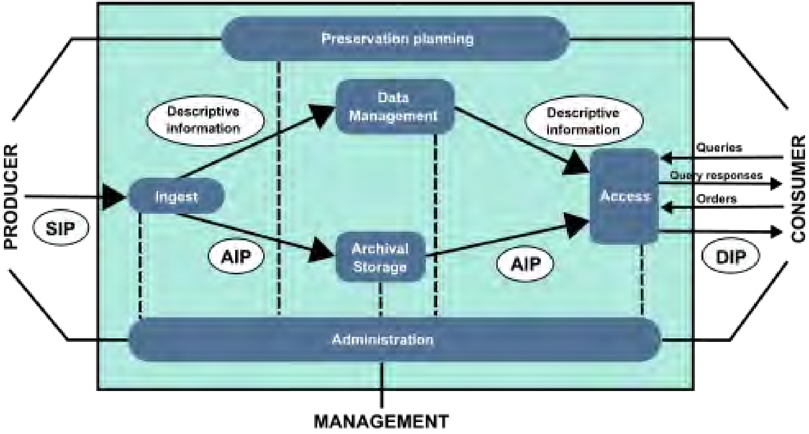
On the producer or creator side:

Computational processing is used by bureaucracies and research:

- to create/gather and analyze data
- in ways that should but may not always generate records for evidentiary, accountability and transparency purposes
- to manage active data and records

Archivists and records managers may be responsible for ensuring that:

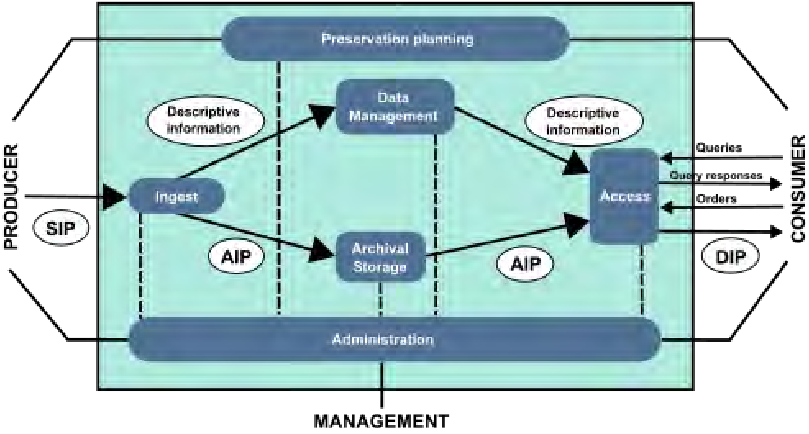
- accountable and transparent records are created and are archivable or disposable
- the algorithms used to produce, process and dispose of the records are also accountable, transparent, accurate and ethical



On the consumer or user side:

- Computational processing is being used in the digital humanities and STEM fields to collate, cross-compile and analyze digitized and born-digital archival materials in new ways in order to gain new insights

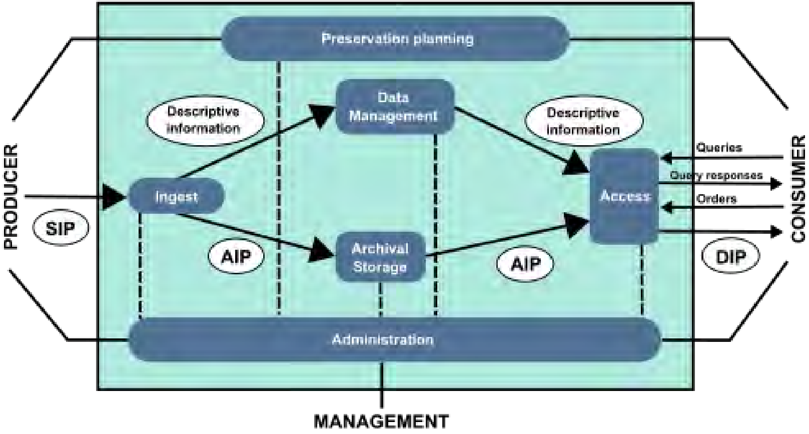
- Archivists have to prepare digital archival content for end-user computational processing, e.g.,
    - by ensuring archival content can be compiled, manipulated and curated at a very granular level
    - by adding descriptive metadata that supports new kinds of disciplinary research questions



During archival appraisal and ingest:

- Human appraisal of digitally-born records is becoming increasingly impossible due to massive volume, complexity and contingency, e.g., digital communications such as email and social media; networked and Cloud-based recordkeeping

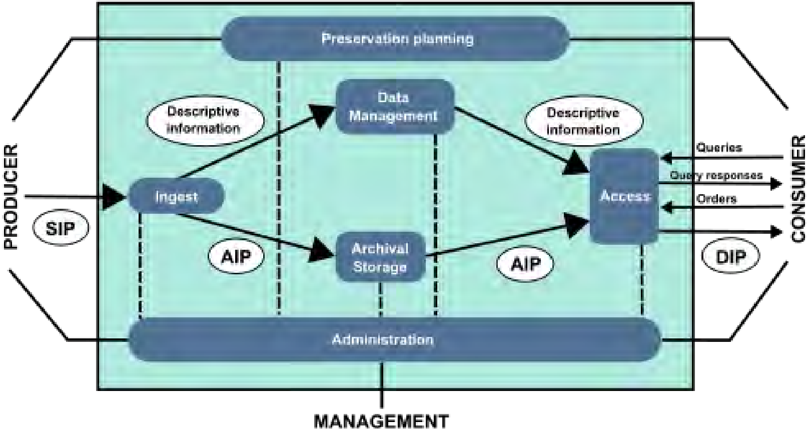
- Archivists will need to employ computational analysis to identify, ingest, and secure relationships between records and their components



During archival preservation:

- Manual preservation of increasing volumes of digital archival content does not scale

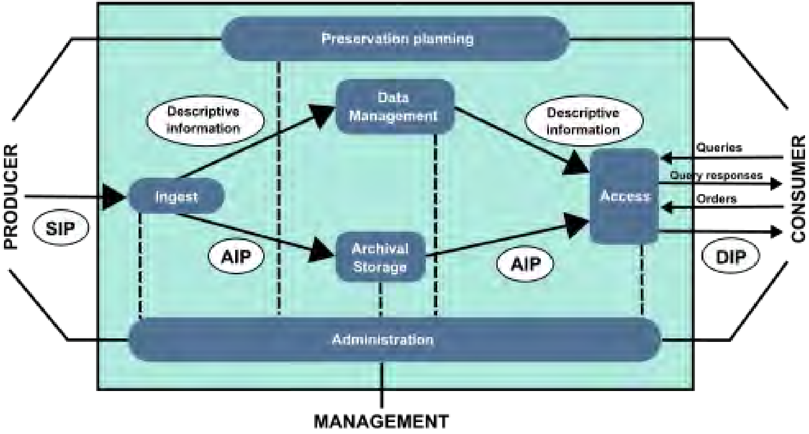
- Archivists will need to employ computational approaches in activities such as integrity checking, regular migration processes and allocating and tracking storage



During archival processing:

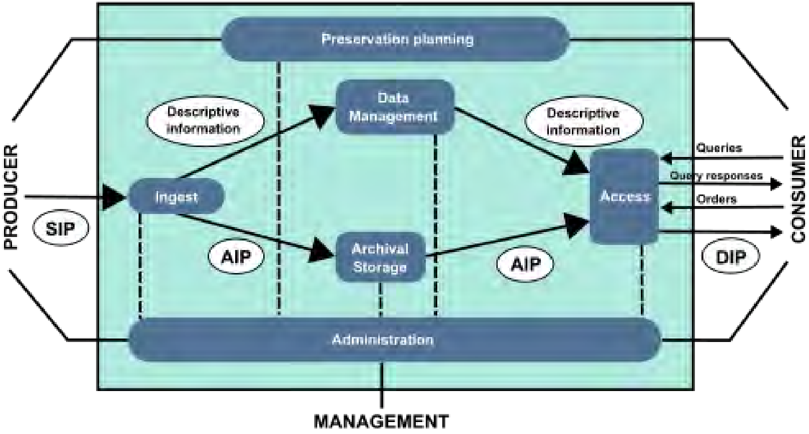
Archivists will use computational processes to:

- automate aspects of description, maintain administrative metadata, surface previously hidden aspects of collections, identify and explain anomalies, and curate collections to support specific researcher/consumer needs



At the access interface:

- Archivists will use computational processes to:
  - establish consumer identity and privileges regarding collection access
  - search for, retrieve and package content that matches consumer queries and specifications
  - ensure privacy and security conditions governing dissemination and access of collections are met



Foundational Paper on Computational Archival Science (CAS): Apr. 2018, Marciano et al.

"Archival records and training in the Age of Big Data"

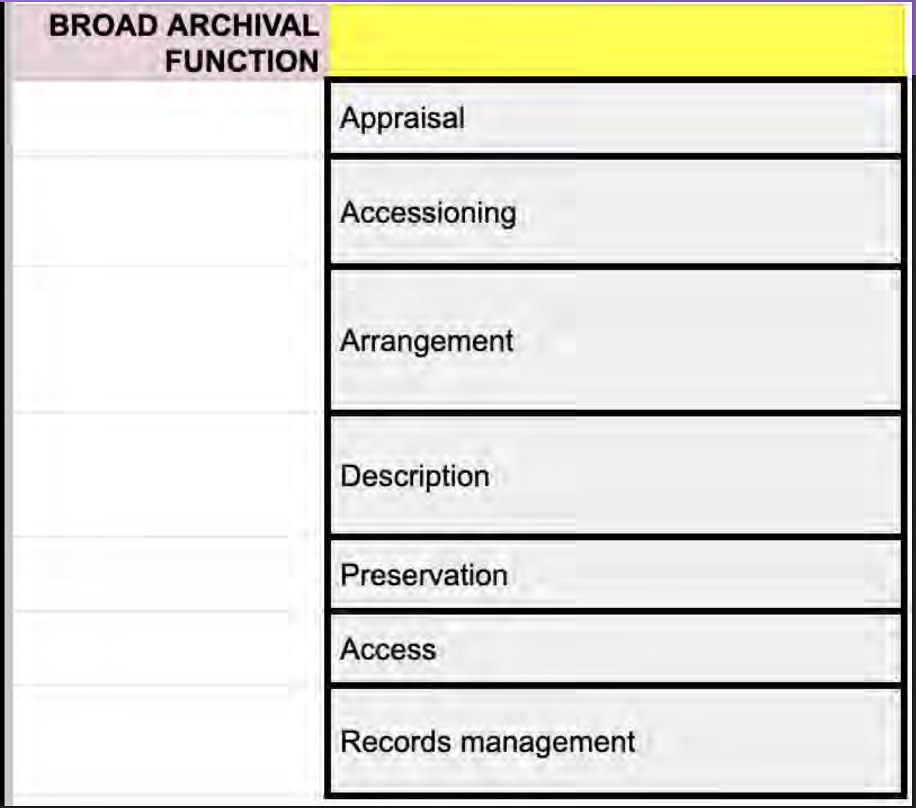
In *Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education.*

https://ai-collaboratory.net/wp-content/uploads/2020/10/Marciano-et-al-Archival-Records-and-Training-in-the-Age-of-Big-Data-final.pdf

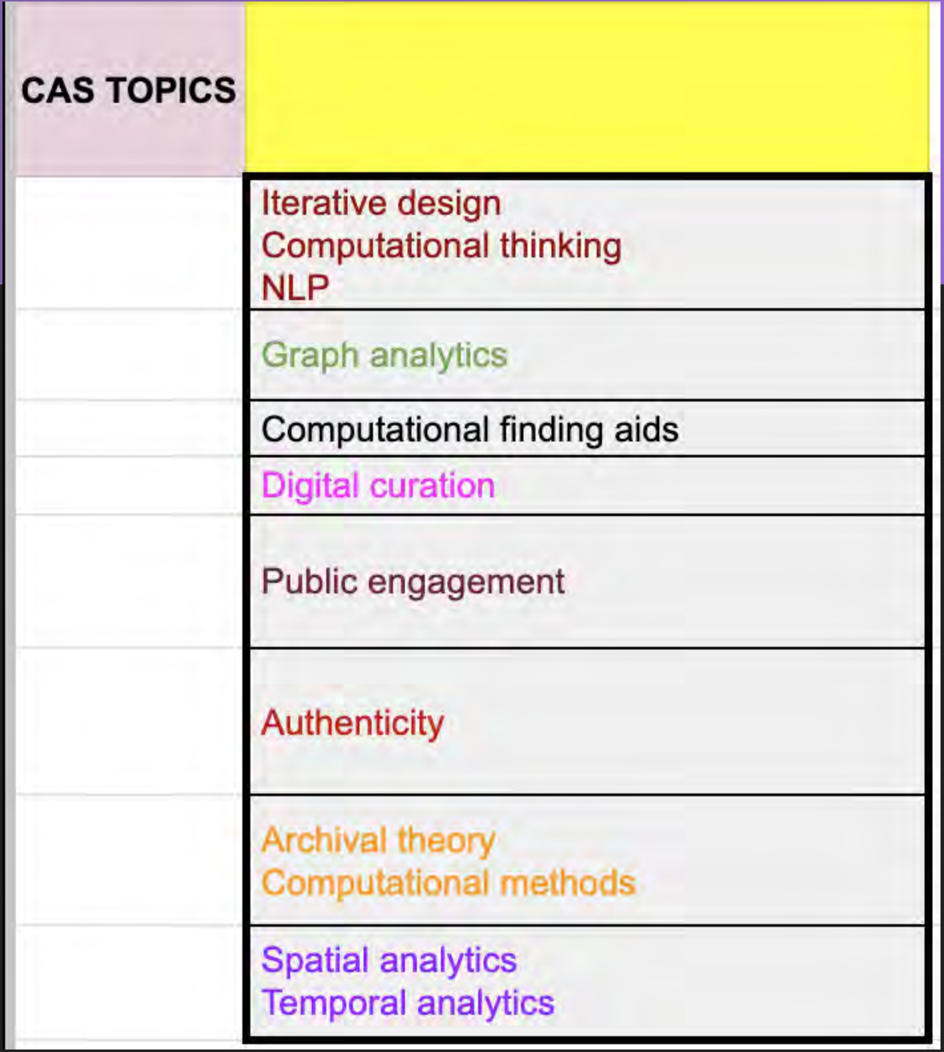
Eight Case Studies w. examples of interdisciplinary efforts to address the changing context of recordkeeping and character of records:

1. **Evolutionary prototyping** and **computational linguistics**,
2. **Graph analytics, digital humanities** and **archival representation**,
3. **Computational finding aids,**
4. **Digital curation**,
5. **Public engagement** with (archival) content,
6. **Authenticity**,
7. Confluences between archival theory and computational methods: **cyberinfrastructure and the Records Continuum**,
8. **Spatial and temporal analytics**.

Each of the case studies concludes with a ***"Takeaways for CAS/MLS Education"*** statement.



BROAD ARCHIVAL FUNCTION	
	Appraisal
	Accessioning
	Arrangement
	Description
	Preservation
	Access
	Records management



CAS TOPICS	
	Iterative design Computational thinking NLP
	Graph analytics
	Computational finding aids
	Digital curation
	Public engagement
	Authenticity
	Archival theory Computational methods
	Spatial analytics Temporal analytics

What is CAS?

GOAL: **Explore computational treatments of archival and cultural content**

GOOGLE GROUP: computational-archival-science@googlegroups.com

Working Definition of Computational Archival Science:

(R. Marciano et al. 2016, amended by N. Payne in 2018)

- A transdisciplinary field grounded in archival, information, and computational science that is concerned with the application of computational methods & resources, design patterns, sociotechnical constructs, and human-technology interaction, to large-scale *(big data)*:
  - records/archives processing, analysis, storage, long-term preservation, and access problems,
- with the aim of improving and optimizing efficiency, authenticity, truthfulness, provenance, productivity, computation, information structure & design, precision & human technology interaction in support of:
  - acquisition, appraisal, arrangement & description, preservation, communication, transmission, analysis &access decisions.

CAS PORTAL: https://ai-collaboratory.net/cas/

* Workshops:

- 50+ workshops since 2016
- 9 CAS @ IEEE Big Data Conf. w. 150+ papers

Lessons learned from:

- **CAS#1**: 2016 in Washington, DC
- **CAS#2**: 2017 in Boston
- **CAS#3**: 2018 in Seattle
- **CAS#4**: 2019 in LA
- **CAS#5**: 2020 in Atlanta
- **CAS#6**: 2021 in Orlando
- **CAS#7**: 2022 in Osaka, Japan
- **CAS#8**: 2023 in Sorrento, Italy
- **CAS#9**: 2024 in Washington, DC

* Presentations

* Publications

* Infrastructure

https://ai-collaboratory.net/cas/cas-workshops/2024-9th-cas-workshop/

Welcome!

2024 IEEE International Conference on Big Data (IEEE BigData 2024)

Dec 15-18, 2024 @ Washington DC, USA

IEEE Big Data 2024: CAS #9

https://ai-collaboratory.net/cas/cas-workshops/2024-9th-cas-workshop/

- **Mon., Nov. 4, 2024 (final):** Due date for full workshop papers submission
- **Fri., Nov. 15, 2024:** Notification of paper acceptance to authors
- **Wed., Nov. 20, 2024 (hard deadline):** Camera-ready of accepted papers
- **Tue., Dec. 17, 2024:** Day-long CAS workshop (in person) in Washington DC, USA

RESEARCH TOPICS COVERED:

- **Application of analytics to archival material**, including AI, ML, text- & data-mining, sentiment analysis, network analysis.
- **Analytics in support of archival processing**, including e-discovery, identification of personal information, appraisal, arrangement and description.
- **Scalable services for archives**, including identification, preservation, metadata generation, integrity checking, normalization, reconciliation, linked data, entity extraction, anonymization and reduction.
- **New forms of archives**, including Web, social media, audiovisual archives, and blockchain.
- **Cyber-infrastructures for archive-based research** and for development and hosting of collections
- **Big data and archival theory and practice**
- **Digital curation and preservation**
- **Crowd-sourcing** and archives
- **Big data and the construction of memory and identity**
- **Specific big data technologies** (e.g. NoSQL databases) and their applications
- **Corpora and reference collections** of big archival data
- **Linked data** and archives
- **Big data and provenance**
- **Constructing big data research objects** from archives
- **Legal and ethical issues** in big data archives

Records Management Journal (Emerald Publishing)

Special 2020 Issue on:

Disruptive technologies for archives & records management and records professionals

Editors:

Julie McLeod, *Northumbria University, UK*
Richard Marciano, *University of Maryland, USA*

Summer 2020: Volume 30, Issue 2 & Issue 3
https://www.emerald.com/insight/content/doi/10.1108/RMJ-07-2020-057/full/html
https://www.emerald.com/insight/publication/issn/0956-5698/vol/30/iss/3

- Algorithm produced records
- Explainable Artificial Intelligence
- Natural Language Processing
- Automated Appraisal
- Internet-of-Things in the Archives
- Managing IoT-data for Gov. Agencies
- Collaboration between AI and Archival Science
- Preserving Virtual Reality
- Record Linking
- Mapping Archival Catalogs from Trees to Networks
- Blockchain and Records Management
- A Code of Ethics for the Digital Age

ACM Journal on Computing and Cultural Heritage (JOCCH)

Special 2022 Issue on:

Computational Archival Science (CAS)

Guest Editors:

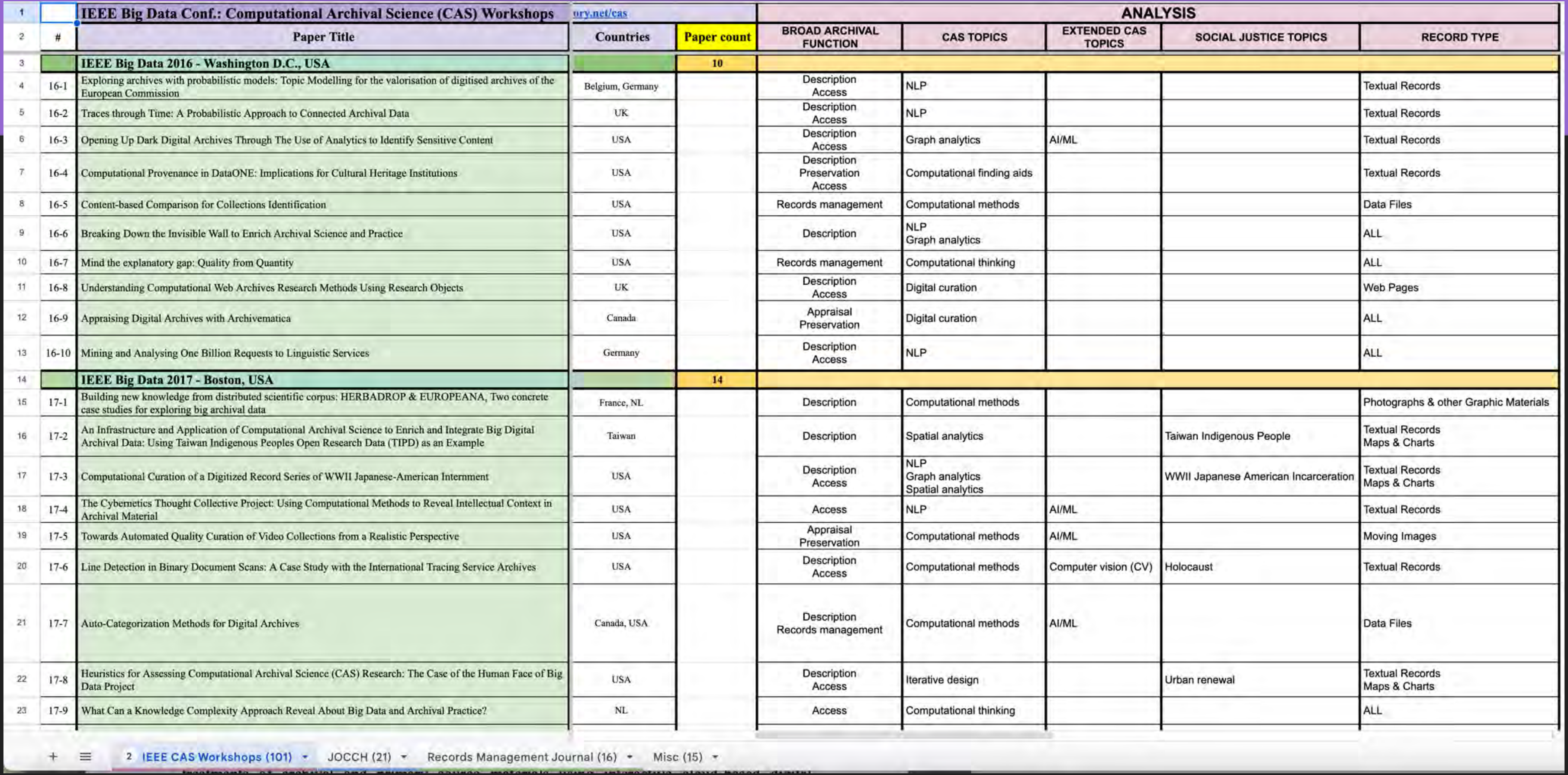
Mark Hedges, *King's College London, UK*
Eirini Goudarouli, *The National Archives, UK*
Richard Marciano, *University of Maryland, USA*

Vol. 13, Issue 1 (Feb. 2022) – Issue 3 (Sep. 2022)

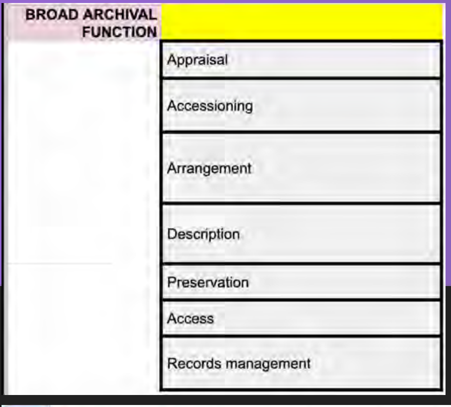
https://ai-collaboratory.net/2020/05/21/jocch-cas_call_for_papers/
https://dl.acm.org/toc/jocch/2022/15/1
https://dl.acm.org/toc/jocch/2022/15/3

Compendium of Core Computational Archival Science (CAS) Papers

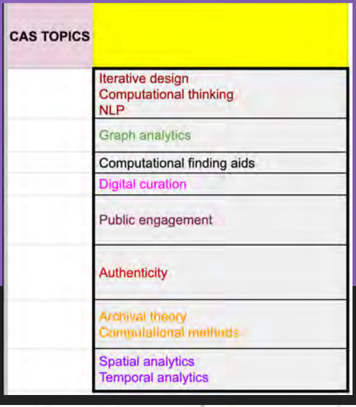
https://docs.google.com/spreadsheets/d/1oCVCWpik_zjdilih9iXh2KITZ7LIOVIt/edit?usp=sharing&ouid=105652788602997060030&rtpof=tr ue&sd=true



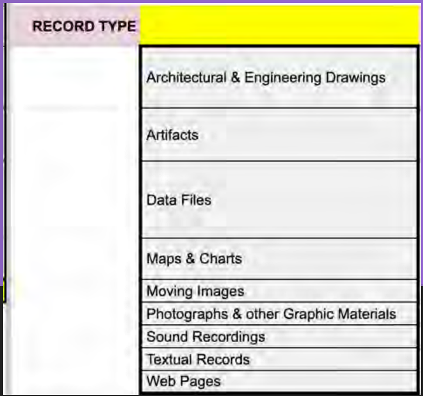
1		IEEE Big Data Conf.: Computational Archival Science (CAS) Workshops	ory.net/cas		ANALYSIS				
2	#	Paper Title	Countries	Paper count	BROAD ARCHIVAL FUNCTION	CAS TOPICS	EXTENDED CAS TOPICS	SOCIAL JUSTICE TOPICS	RECORD TYPE
3		IEEE Big Data 2016 - Washington D.C., USA		10					
4	16-1	Exploring archives with probabilistic models: Topic Modelling for the valorisation of digitised archives of the European Commission	Belgium, Germany		Description Access	NLP			Textual Records
5	16-2	Traces through Time: A Probabilistic Approach to Connected Archival Data	UK		Description Access	NLP			Textual Records
6	16-3	Opening Up Dark Digital Archives Through The Use of Analytics to Identify Sensitive Content	USA		Description Access	Graph analytics	AI/ML		Textual Records
7	16-4	Computational Provenance in DataONE: Implications for Cultural Heritage Institutions	USA		Description Preservation Access	Computational finding aids			Textual Records
8	16-5	Content-based Comparison for Collections Identification	USA		Records management	Computational methods			Data Files
9	16-6	Breaking Down the Invisible Wall to Enrich Archival Science and Practice	USA		Description	NLP Graph analytics			ALL
10	16-7	Mind the explanatory gap: Quality from Quantity	USA		Records management	Computational thinking			ALL
11	16-8	Understanding Computational Web Archives Research Methods Using Research Objects	UK		Description Access	Digital curation			Web Pages
12	16-9	Appraising Digital Archives with Archivematica	Canada		Appraisal Preservation	Digital curation			ALL
13	16-10	Mining and Analysing One Billion Requests to Linguistic Services	Germany		Description Access	NLP			ALL
14		IEEE Big Data 2017 - Boston, USA		14					
15	17-1	Building new knowledge from distributed scientific corpus: HERBADROP & EUROPEANA, Two concrete case studies for exploring big archival data	France, NL		Description	Computational methods			Photographs & other Graphic Materials
16	17-2	An Infrastructure and Application of Computational Archival Science to Enrich and Integrate Big Digital Archival Data: Using Taiwan Indigenous Peoples Open Research Data (TIPD) as an Example	Taiwan		Description	Spatial analytics		Taiwan Indigenous People	Textual Records Maps & Charts
17	17-3	Computational Curation of a Digitized Record Series of WWII Japanese-American Internment	USA		Description Access	NLP Graph analytics Spatial analytics		WWII Japanese American Incarceration	Textual Records Maps & Charts
18	17-4	The Cybernetics Thought Collective Project: Using Computational Methods to Reveal Intellectual Context in Archival Material	USA		Access	NLP	AI/ML		Textual Records
19	17-5	Towards Automated Quality Curation of Video Collections from a Realistic Perspective	USA		Appraisal Preservation	Computational methods	AI/ML		Moving Images
20	17-6	Line Detection in Binary Document Scans: A Case Study with the International Tracing Service Archives	USA		Description Access	Computational methods	Computer vision (CV)	Holocaust	Textual Records
21	17-7	Auto-Categorization Methods for Digital Archives	Canada, USA		Description Records management	Computational methods	AI/ML		Data Files
22	17-8	Heuristics for Assessing Computational Archival Science (CAS) Research: The Case of the Human Face of Big Data Project	USA		Description Access	Iterative design		Urban renewal	Textual Records Maps & Charts
23	17-9	What Can a Knowledge Complexity Approach Reveal About Big Data and Archival Practice?	NL		Access	Computational thinking			ALL



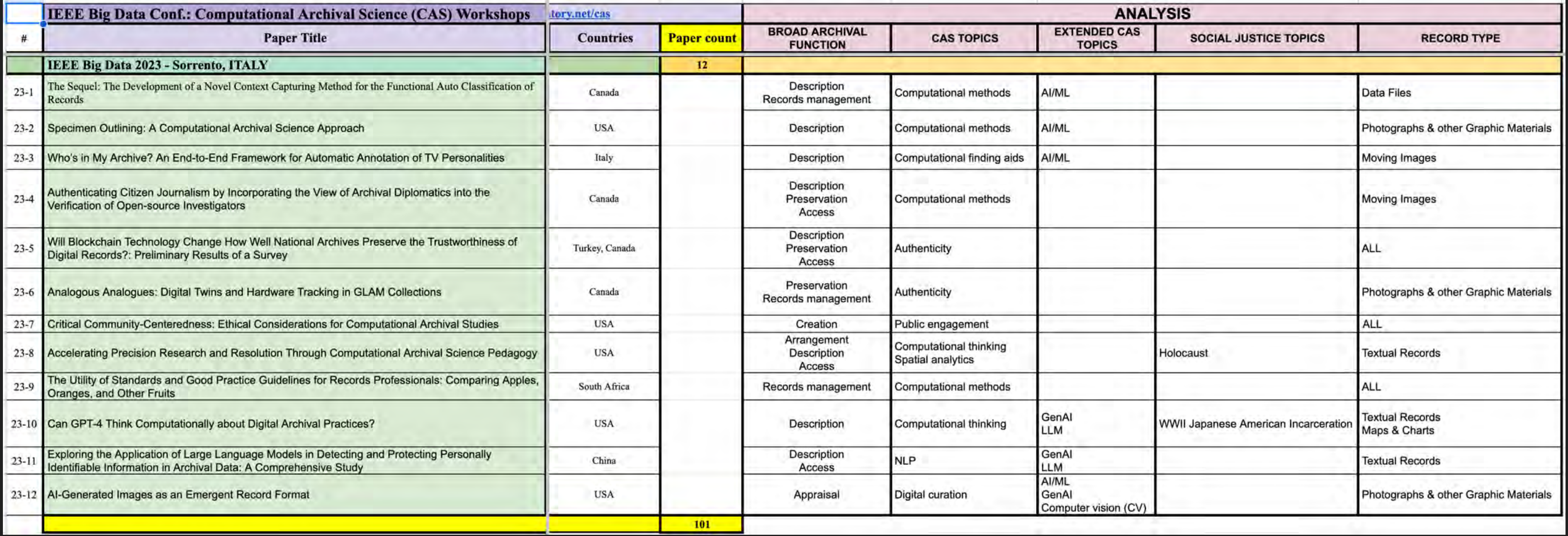
BROAD ARCHIVAL FUNCTION	
	Appraisal
	Accessioning
	Arrangement
	Description
	Preservation
	Access
	Records management



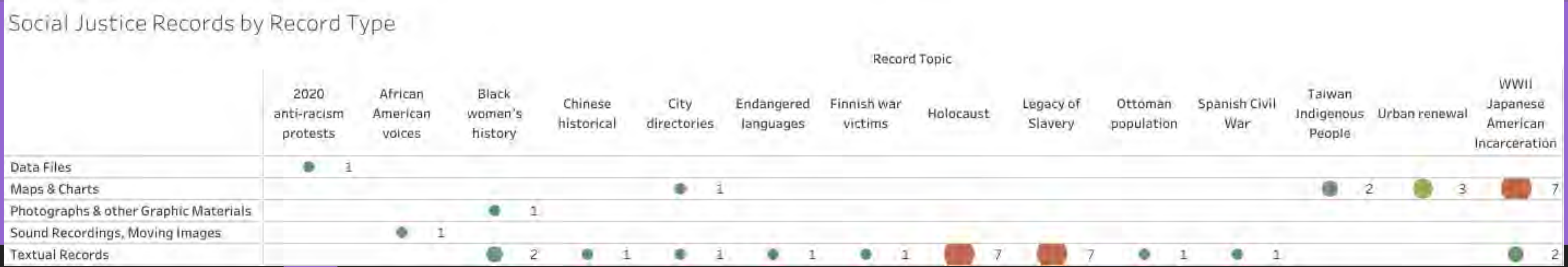
CAS TOPICS	
	Iterative design Computational thinking NLP
	Graph analytics
	Computational finding aids
	Digital curation
	Public engagement
	Authenticity
	Archival theory Computational methods
	Spatial analytics Temporal analytics



RECORD TYPE	
	Architectural & Engineering Drawings
	Artifacts
	Data Files
	Maps & Charts
	Moving Images
	Photographs & other Graphic Materials
	Sound Recordings
	Textual Records
	Web Pages

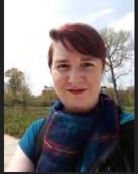


	IEEE Big Data Conf.: Computational Archival Science (CAS) Workshops	tory.net/cas		ANALYSIS				
#	Paper Title	Countries	Paper count	BROAD ARCHIVAL FUNCTION	CAS TOPICS	EXTENDED CAS TOPICS	SOCIAL JUSTICE TOPICS	RECORD TYPE
	IEEE Big Data 2023 - Sorrento, ITALY		12					
23-1	The Sequel: The Development of a Novel Context Capturing Method for the Functional Auto Classification of Records	Canada		Description Records management	Computational methods	AI/ML		Data Files
23-2	Specimen Outlining: A Computational Archival Science Approach	USA		Description	Computational methods	AI/ML		Photographs & other Graphic Materials
23-3	Who's in My Archive? An End-to-End Framework for Automatic Annotation of TV Personalities	Italy		Description	Computational finding aids	AI/ML		Moving Images
23-4	Authenticating Citizen Journalism by Incorporating the View of Archival Diplomatics into the Verification of Open-source Investigators	Canada		Description Preservation Access	Computational methods			Moving Images
23-5	Will Blockchain Technology Change How Well National Archives Preserve the Trustworthiness of Digital Records?: Preliminary Results of a Survey	Turkey, Canada		Description Preservation Access	Authenticity			ALL
23-6	Analogous Analogues: Digital Twins and Hardware Tracking in GLAM Collections	Canada		Preservation Records management	Authenticity			Photographs & other Graphic Materials
23-7	Critical Community-Centeredness: Ethical Considerations for Computational Archival Studies	USA		Creation	Public engagement			ALL
23-8	Accelerating Precision Research and Resolution Through Computational Archival Science Pedagogy	USA		Arrangement Description Access	Computational thinking Spatial analytics		Holocaust	Textual Records
23-9	The Utility of Standards and Good Practice Guidelines for Records Professionals: Comparing Apples, Oranges, and Other Fruits	South Africa		Records management	Computational methods			ALL
23-10	Can GPT-4 Think Computationally about Digital Archival Practices?	USA		Description	Computational thinking	GenAI LLM	WWII Japanese American Incarceration	Textual Records Maps & Charts
23-11	Exploring the Application of Large Language Models in Detecting and Protecting Personally Identifiable Information in Archival Data: A Comprehensive Study	China		Description Access	NLP	GenAI LLM		Textual Records
23-12	AI-Generated Images as an Emergent Record Format	USA		Appraisal	Digital curation	AI/ML GenAI Computer vision (CV)		Photographs & other Graphic Materials
			101					



Social Justice Records by Record Type

	Record Topic													
	2020 anti-racism protests	African American voices	Black women's history	Chinese historical	City directories	Endangered languages	Finnish war victims	Holocaust	Legacy of Slavery	Ottoman population	Spanish Civil War	Taiwan Indigenous People	Urban renewal	WWII Japanese American Incarceration
Data Files	1													
Maps & Charts					1							2	3	7
Photographs & other Graphic Materials			1											
Sound Recordings, Moving Images		1												
Textual Records			2	1	1	1	1	7	7	1	1			2

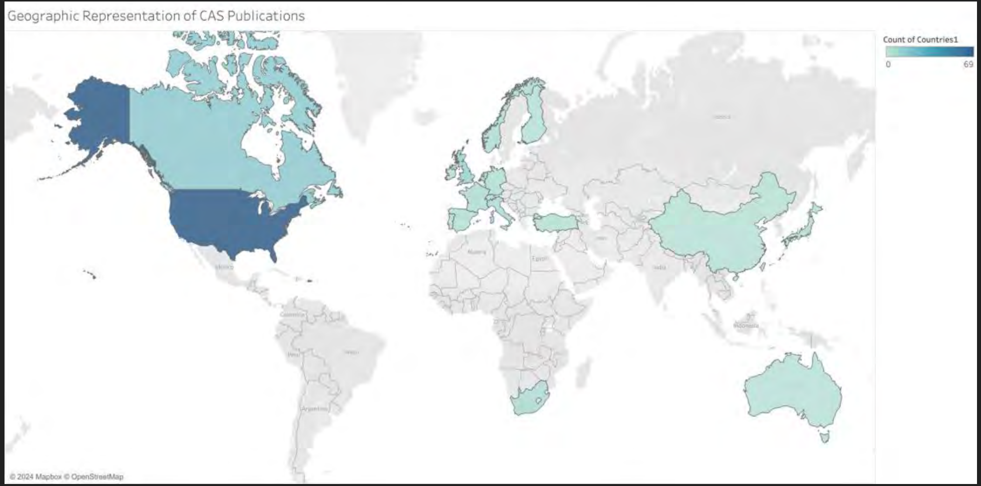


Jennifer Proctor

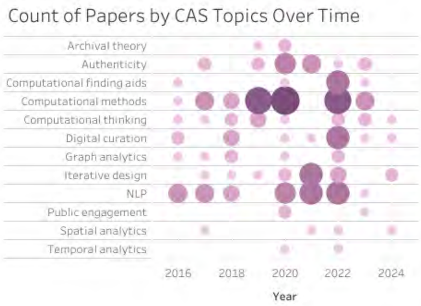
DublinCore Webinar: Oct. 19, 2023

AI & NLP for Open-Source Archival Linked Data Workflows

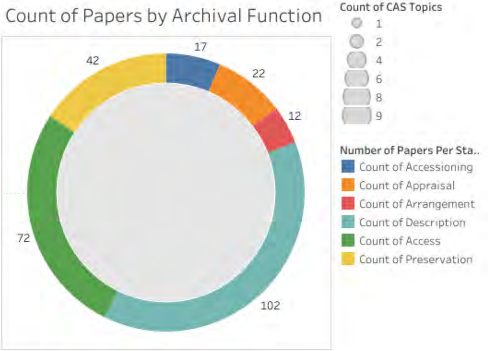
https://www.dublincore.org/webinars/2023/ai-nlp-archival-linked-data/



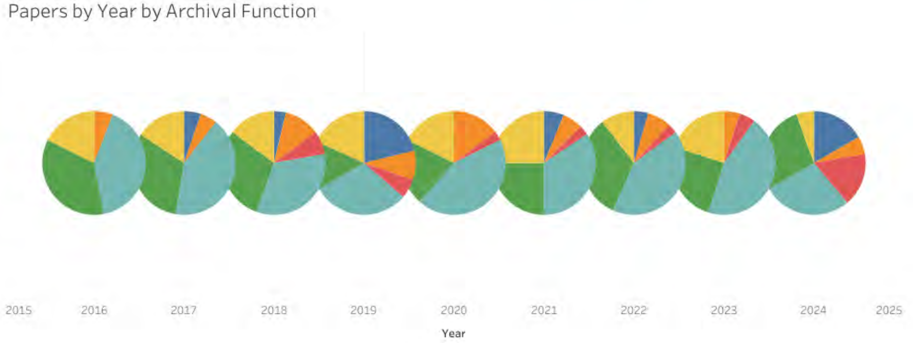
Geographic Representation of CAS Publications



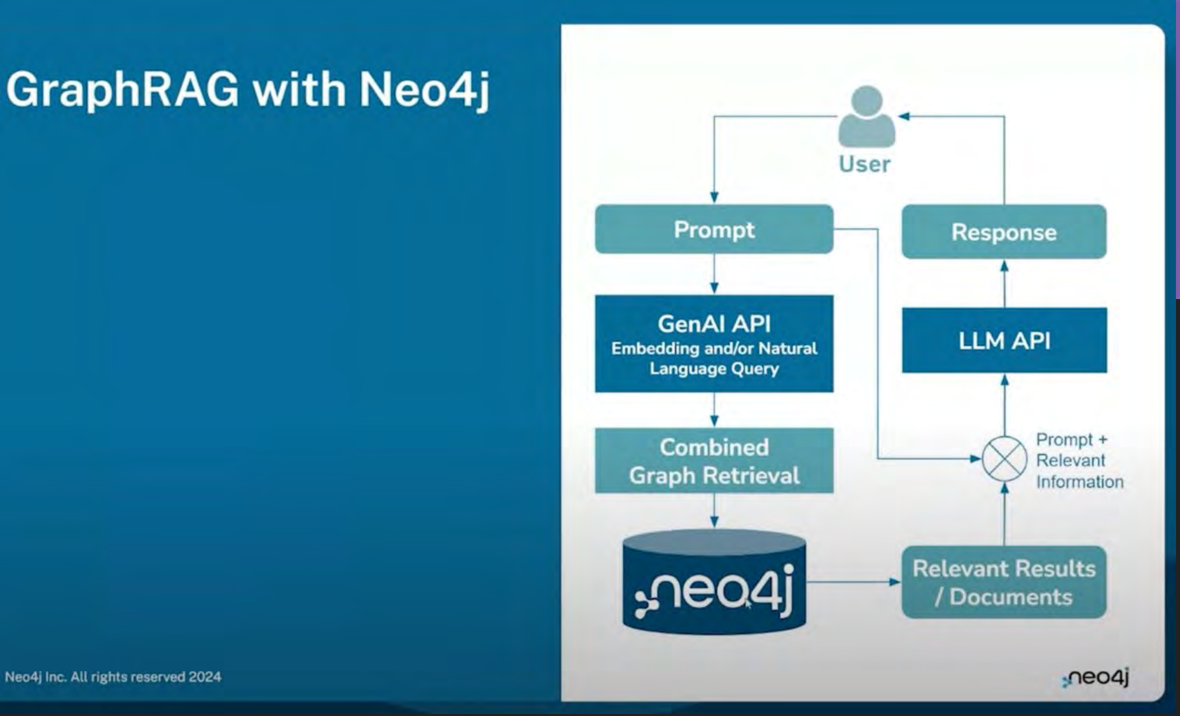
Count of Papers by CAS Topics Over Time



Count of Papers by Archival Function



Papers by Year by Archival Function



GraphRAG with Neo4j

User
Prompt
Response
GenAI API
Embedding and/or Natural Language Query
LLM API
Combined Graph Retrieval
Prompt + Relevant Information
neo4j
Relevant Results / Documents
neo4j

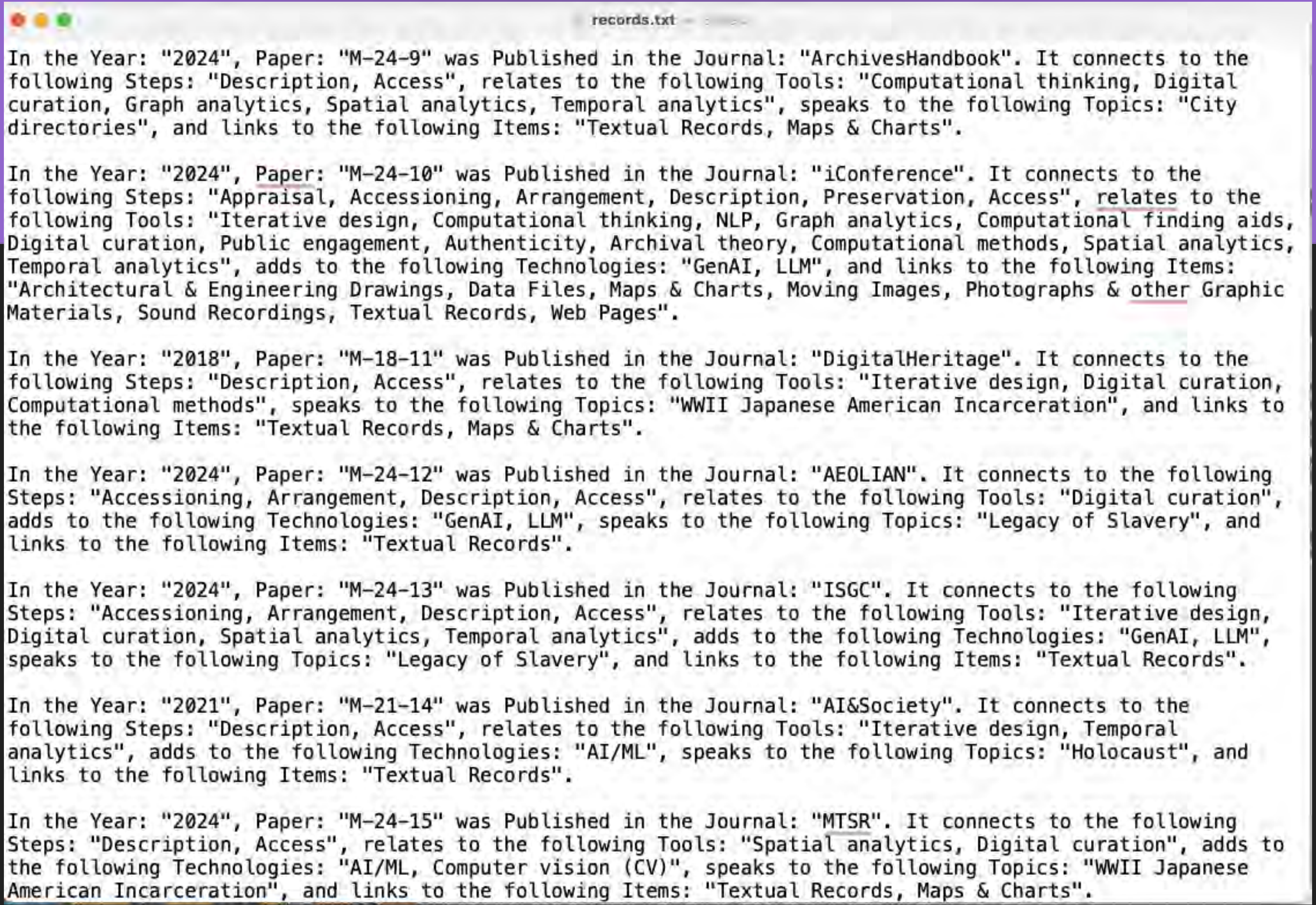
Neo4j Inc. All rights reserved 2024

Neo4j LLM Knowledge Graph Builder
Allison Cossette
https://www.youtube.com/watch?v=UYJbG3p68NM

GraphRAG (Retrieval Augmented Generation with Graphs):

Merges Knowledge Graphs ”with context, structure & trust”.

Instead of relying only on text chunk searches (using Vector DBs), it uses graph queries to pull relevant, connected data. The claim is that it helps with “Explainable AI”: every result comes with an audit trail, helping with trust and transparency.



In the Year: "2024", Paper: "M-24-9" was Published in the Journal: "ArchivesHandbook". It connects to the following Steps: "Description, Access", relates to the following Tools: "Computational thinking, Digital curation, Graph analytics, Spatial analytics, Temporal analytics", speaks to the following Topics: "City directories", and links to the following Items: "Textual Records, Maps & Charts".

In the Year: "2024", Paper: "M-24-10" was Published in the Journal: "iConference". It connects to the following Steps: "Appraisal, Accessioning, Arrangement, Description, Preservation, Access", relates to the following Tools: "Iterative design, Computational thinking, NLP, Graph analytics, Computational finding aids, Digital curation, Public engagement, Authenticity, Archival theory, Computational methods, Spatial analytics, Temporal analytics", adds to the following Technologies: "GenAI, LLM", and links to the following Items: "Architectural & Engineering Drawings, Data Files, Maps & Charts, Moving Images, Photographs & other Graphic Materials, Sound Recordings, Textual Records, Web Pages".

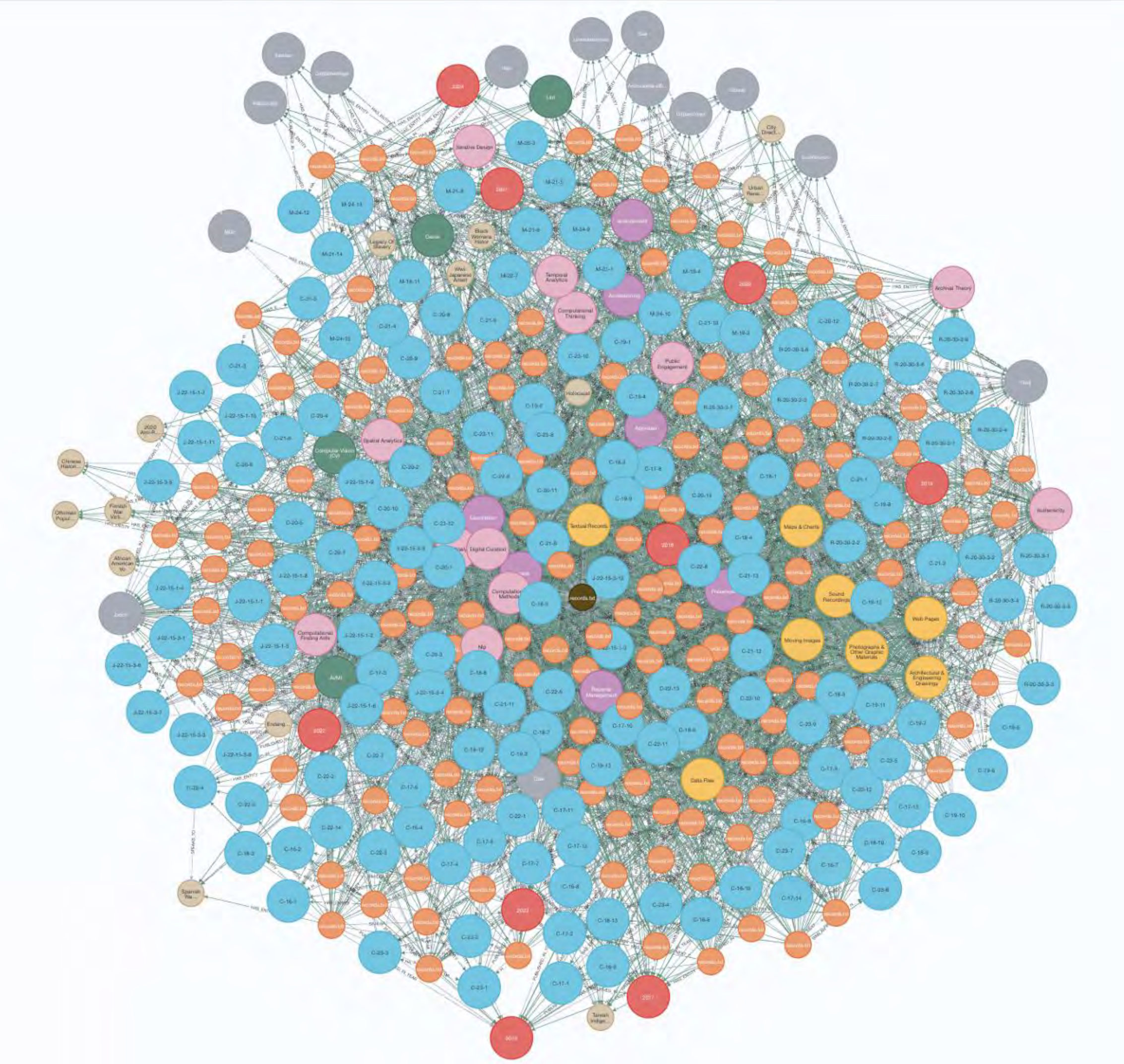
In the Year: "2018", Paper: "M-18-11" was Published in the Journal: "DigitalHeritage". It connects to the following Steps: "Description, Access", relates to the following Tools: "Iterative design, Digital curation, Computational methods", speaks to the following Topics: "WWII Japanese American Incarceration", and links to the following Items: "Textual Records, Maps & Charts".

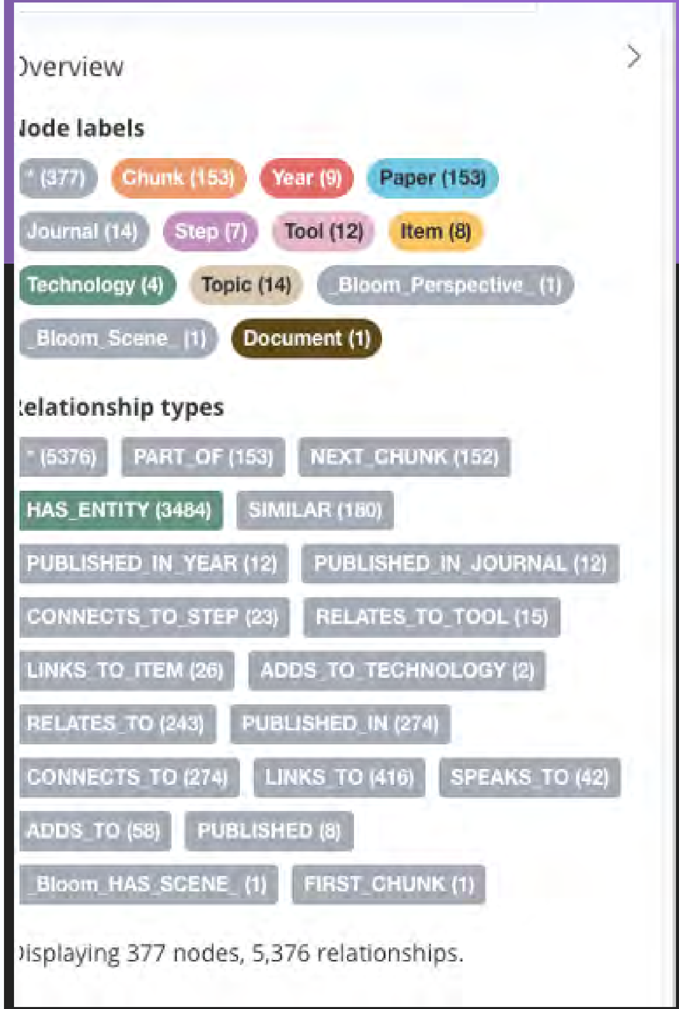
In the Year: "2024", Paper: "M-24-12" was Published in the Journal: "AEOLIAN". It connects to the following Steps: "Accessioning, Arrangement, Description, Access", relates to the following Tools: "Digital curation", adds to the following Technologies: "GenAI, LLM", speaks to the following Topics: "Legacy of Slavery", and links to the following Items: "Textual Records".

In the Year: "2024", Paper: "M-24-13" was Published in the Journal: "ISGC". It connects to the following Steps: "Accessioning, Arrangement, Description, Access", relates to the following Tools: "Iterative design, Digital curation, Spatial analytics, Temporal analytics", adds to the following Technologies: "GenAI, LLM", speaks to the following Topics: "Legacy of Slavery", and links to the following Items: "Textual Records".

In the Year: "2021", Paper: "M-21-14" was Published in the Journal: "AI&Society". It connects to the following Steps: "Description, Access", relates to the following Tools: "Iterative design, Temporal analytics", adds to the following Technologies: "AI/ML", speaks to the following Topics: "Holocaust", and links to the following Items: "Textual Records".

In the Year: "2024", Paper: "M-24-15" was Published in the Journal: "MTSR". It connects to the following Steps: "Description, Access", relates to the following Tools: "Spatial analytics, Digital curation", adds to the following Technologies: "AI/ML, Computer vision (CV)", speaks to the following Topics: "WWII Japanese American Incarceration", and links to the following Items: "Textual Records, Maps & Charts".





verview

ode labels

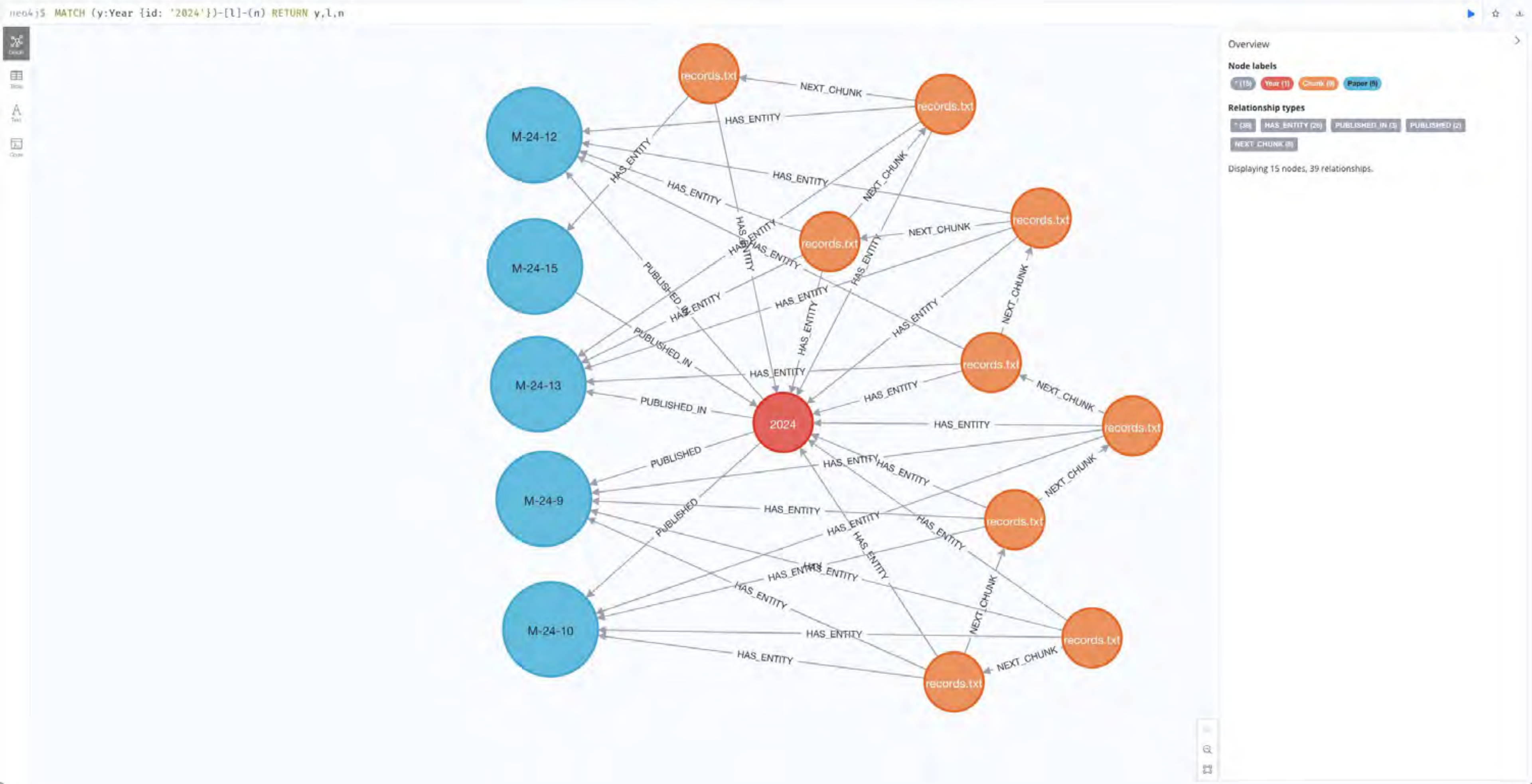
* (377) Chunk (153) Year (9) Paper (153) Journal (14) Step (7) Tool (12) Item (8) Technology (4) Topic (14) _Bloom_Perspective_ (1) _Bloom_Scene_ (1) Document (1)

elationship types

* (5376) PART_OF (153) NEXT_CHUNK (152) HAS_ENTITY (3484) SIMILAR (180) PUBLISHED_IN_YEAR (12) PUBLISHED_IN_JOURNAL (12) CONNECTS_TO_STEP (23) RELATES_TO_TOOL (15) LINKS_TO_ITEM (26) ADDS_TO_TECHNOLOGY (2) RELATES_TO (243) PUBLISHED_IN (274) CONNECTS_TO (274) LINKS_TO (416) SPEAKS_TO (42) ADDS_TO (58) PUBLISHED (8) _Bloom_HAS_SCENE_ (1) FIRST_CHUNK (1)

isplaying 377 nodes, 5,376 relationships.

SUBSET THE KG showing only nodes connected to Year ‘2024’:



```
neo4j$ MATCH (y:Year {id: '2024'})-[l]-(n) RETURN y,l,n
```

Overview

Node labels

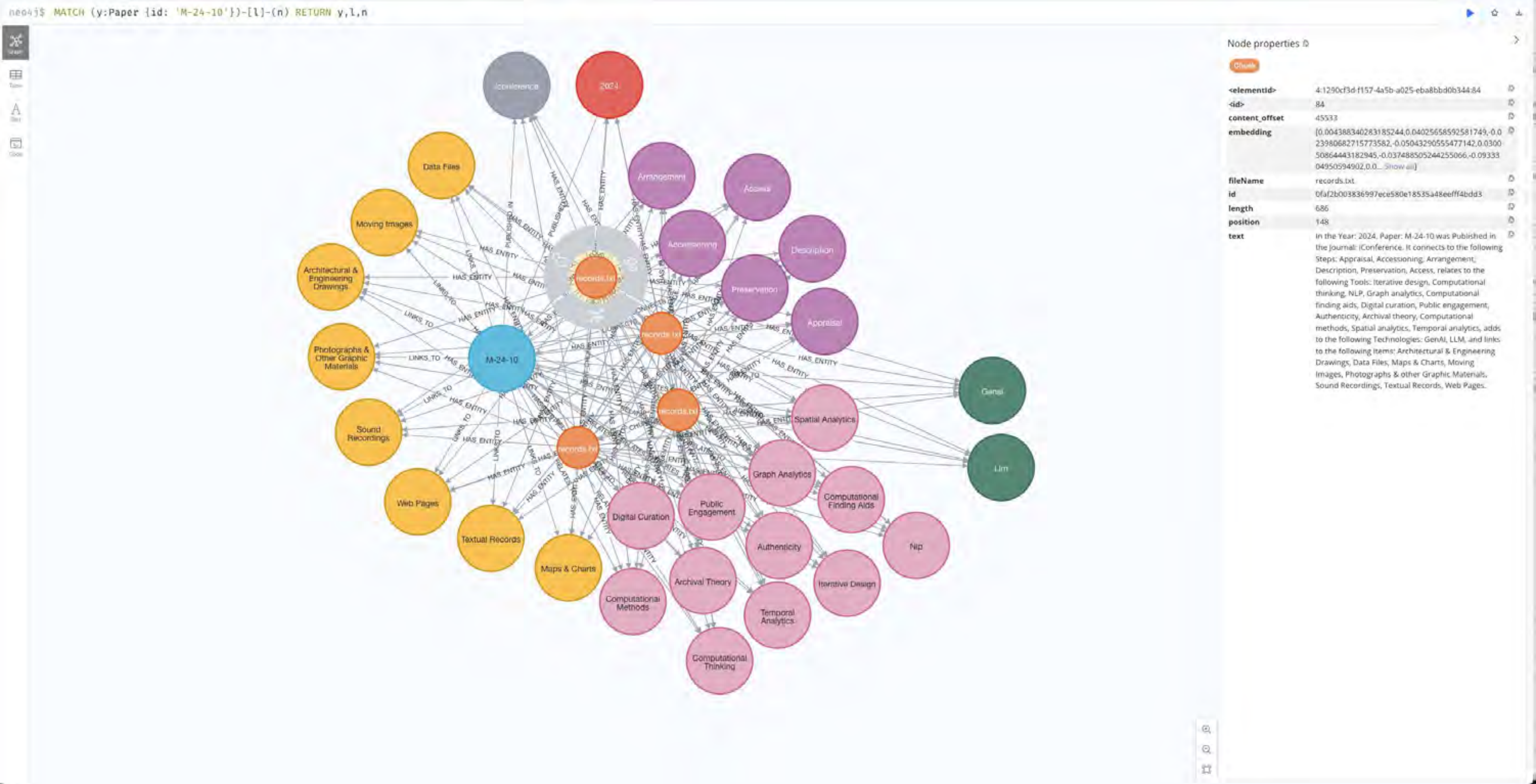
* (15) Year (1) Chunk (9) Paper (5)

Relationship types

* (39) HAS_ENTITY (26) PUBLISHED_IN (3) PUBLISHED (2) NEXT_CHUNK (8)

Displaying 15 nodes, 39 relationships.

SUBSET THE KG showing only nodes connected to Paper ‘M-24-10’:



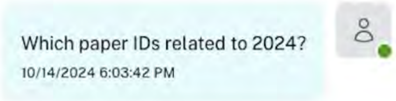
```
neo4j$ MATCH (y:Paper {id: 'M-24-10'})-[l]-(n) RETURN y,l,n
```

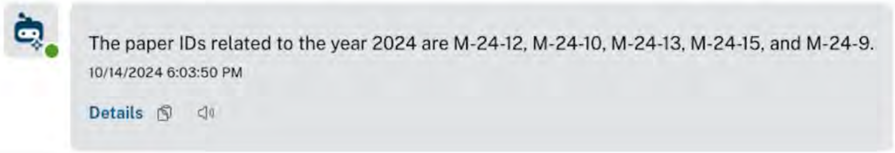
Node properties

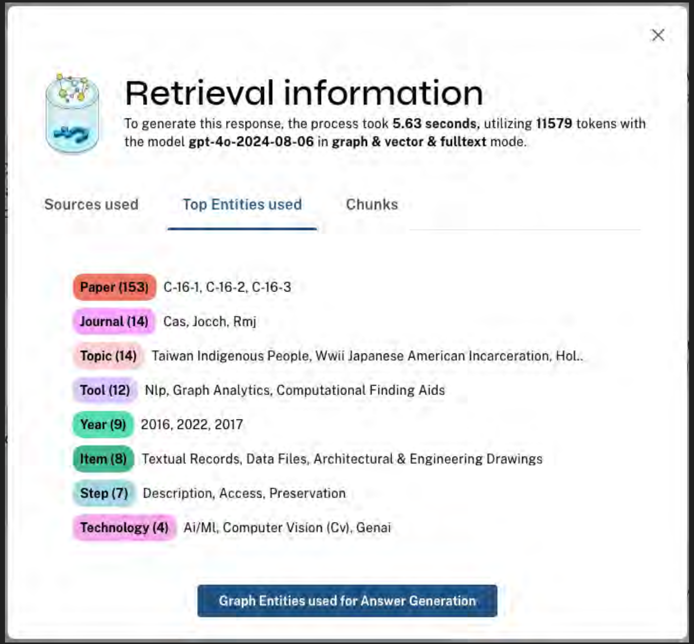
Chunk

<elementid>	4:1290cf3d-f157-4a5b-a025-eba8bbd0b344:84
<id>	84
content_offset	45533
embedding	[0.004388340283185244,0.04025658592581749,-0.023980682715773582,-0.05043290555477142,0.030050864443182945,-0.037488505244255066,-0.0933304950594902,0.0... Show all]
fileName	records.txt
id	0faf2b003836997ece580e18535a48eefff4bdd3
length	686
position	148
text	In the Year: 2024, Paper: M-24-10 was Published in the Journal: iConference. It connects to the following Steps: Appraisal, Accessioning, Arrangement, Description, Preservation, Access, relates to the following Tools: Iterative design, Computational thinking, NLP, Graph analytics, Computational finding aids, Digital curation, Public engagement, Authenticity, Archival theory, Computational methods, Spatial analytics, Temporal analytics, adds to the following Technologies: GenAI, LLM, and links to the following Items: Architectural & Engineering Drawings, Data Files, Maps & Charts, Moving Images, Photographs & other Graphic Materials, Sound Recordings, Textual Records, Web Pages.

GraphRAG using ChatGPT4.o prompts:


Which paper IDs related to 2024?
10/14/2024 6:03:42 PM

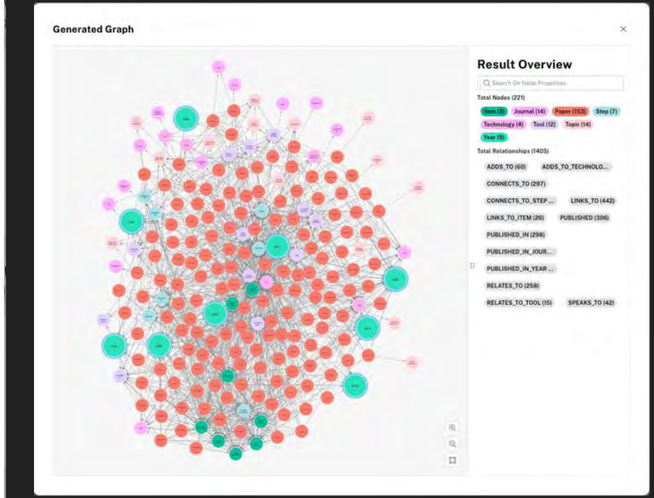

The paper IDs related to the year 2024 are M-24-12, M-24-10, M-24-13, M-24-15, and M-24-9.
10/14/2024 6:03:50 PM


Retrieval information

To generate this response, the process took **5.63 seconds,** utilizing **11579** tokens with the model **gpt-4o-2024-08-06** in **graph & vector & fulltext** mode.

Paper (153)	C-16-1, C-16-2, C-16-3
Journal (14)	Cas, Jocch, Rmj
Topic (14)	Taiwan Indigenous People, Wwii Japanese American Incarceration, Hol..
Tool (12)	Nlp, Graph Analytics, Computational Finding Aids
Year (9)	2016, 2022, 2017
Item (8)	Textual Records, Data Files, Architectural & Engineering Drawings
Step (7)	Description, Access, Preservation
Technology (4)	Ai/Ml, Computer Vision (Cv), Genai

Graph Entities used for Answer Generation


Generated Graph

Result Overview

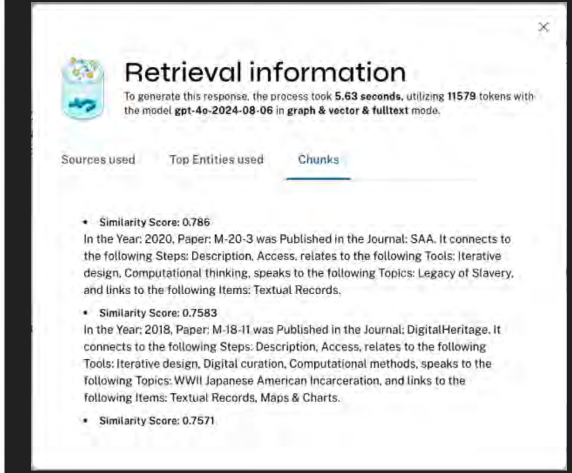
Search On Node Properties

Total Nodes (221)

Item (8) Journal (14) Paper (153) Step (7) Technology (4) Tool (12) Topic (14) Year (9)

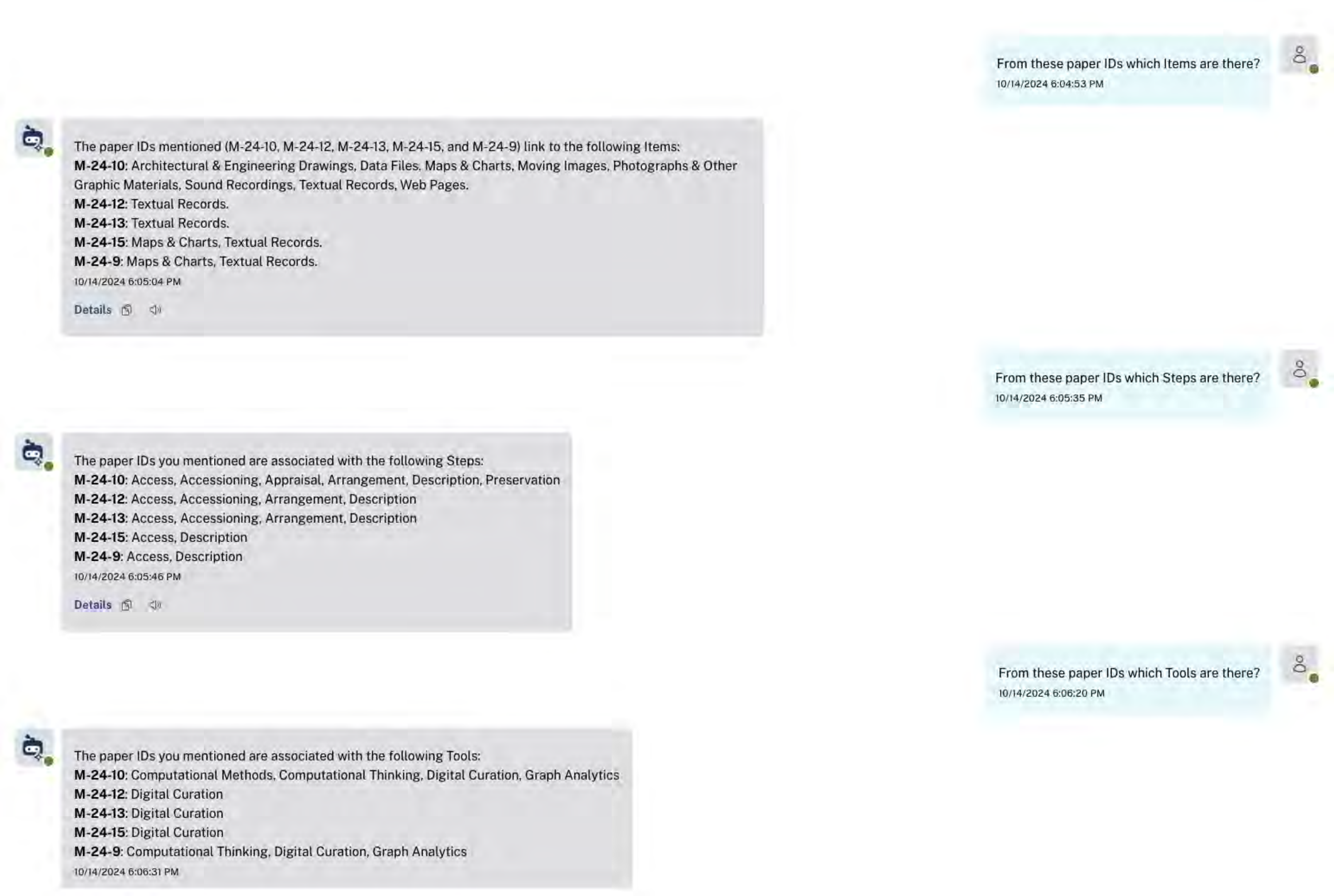
Total Relationships (1405)

ADDS_TO (60) ADDS_TO_TECHNOLO...
CONNECTS_TO (297)
CONNECTS_TO_STEP ... LINKS_TO (442)
LINKS_TO_ITEM (26) PUBLISHED (306)
PUBLISHED_IN (298)
PUBLISHED_IN_JOUR...
PUBLISHED_IN_YEAR ...
RELATES_TO (258)
RELATES_TO_TOOL (15) SPEAKS_TO (42)


Retrieval information

To generate this response, the process took **5.63 seconds,** utilizing **11579** tokens with the model **gpt-4o-2024-08-06** in **graph & vector & fulltext** mode.

- **Similarity Score: 0.786**
In the Year: 2020, Paper: M-20-3 was Published in the Journal: SAA. It connects to the following Steps: Description, Access, relates to the following Tools: Iterative design, Computational thinking, speaks to the following Topics: Legacy of Slavery, and links to the following Items: Textual Records.
- **Similarity Score: 0.7583**
In the Year: 2018, Paper: M-18-11 was Published in the Journal: DigitalHeritage. It connects to the following Steps: Description, Access, relates to the following Tools: Iterative design, Digital curation, Computational methods, speaks to the following Topics: WWII Japanese American Incarceration, and links to the following Items: Textual Records, Maps & Charts.
- **Similarity Score: 0.7571**



From these paper IDs which Items are there?
10/14/2024 6:04:53 PM

The paper IDs mentioned (M-24-10, M-24-12, M-24-13, M-24-15, and M-24-9) link to the following Items:
M-24-10: Architectural & Engineering Drawings, Data Files, Maps & Charts, Moving Images, Photographs & Other Graphic Materials, Sound Recordings, Textual Records, Web Pages.
M-24-12: Textual Records.
M-24-13: Textual Records.
M-24-15: Maps & Charts, Textual Records.
M-24-9: Maps & Charts, Textual Records.
10/14/2024 6:05:04 PM

From these paper IDs which Steps are there?
10/14/2024 6:05:35 PM

The paper IDs you mentioned are associated with the following Steps:
M-24-10: Access, Accessioning, Appraisal, Arrangement, Description, Preservation
M-24-12: Access, Accessioning, Arrangement, Description
M-24-13: Access, Accessioning, Arrangement, Description
M-24-15: Access, Description
M-24-9: Access, Description
10/14/2024 6:05:46 PM

From these paper IDs which Tools are there?
10/14/2024 6:06:20 PM

The paper IDs you mentioned are associated with the following Tools:
M-24-10: Computational Methods, Computational Thinking, Digital Curation, Graph Analytics
M-24-12: Digital Curation
M-24-13: Digital Curation
M-24-15: Digital Curation
M-24-9: Computational Thinking, Digital Curation, Graph Analytics
10/14/2024 6:06:31 PM

Richard Marciano

College of INFORMATION

Google Groups — computational-archival-science@googlegroups.com

Advanced Information Collaboratory (AI) — https://ai-collaboratory.net/

University of Maryland College of Information — marciano@umd.edu

LinkedIn — https://www.linkedin.com/in/richard-marciano/