A Computational Review of the Literature of Computational Archival Science (CAS):

Advancing Archival Theory in the Age of the Digital Tsunami and the Vanishing Box Problem

Jennifer Proctor
Applied Research Laboratory for Intelligence and Security
University of Maryland
College Park, USA
jeproc@umd.edu

Richard Marciano
College of Information
University of Maryland
College Park, USA
marciano@umd.edu

https://ai-collaboratory.net/

CAS: A Survey

Archival Science (CAS) Workshops	ANALYSIS			
Title	BROAD ARCHIVAL FUNCTION	CAS TOPICS	EXTENDED CAS TOPICS	SOCIAL JUSTIC
ring Method for the Functional Auto Classification of	Description Records management	Computational methods	AI/ML	
nce Approach	Description	Computational methods	AI/ML	
r Automatic Annotation of TV Personalities	Description	Computational finding aids	AI/ML	
he View of Archival Diplomatics into the	Description Preservation Access	Computational methods		
nal Archives Preserve the Trustworthiness of	Description Preservation Access	Authenticity		
Tracking in GLAM Collections	Preservation Records management	Authenticity		
rations for Computational Archival Studies	Creation	Public engagement		
rough Computational Archival Science Pedagogy	Arrangement Description Access	Computational thinking Spatial analytics		Holocaust
es for Records Professionals: Comparing Apples,	Records management	Computational methods		
hival Practices?	Description	Computational thinking	GenAI LLM	WWII Japanese Ameri
s in Detecting and Protecting Personally hensive Study	Description Access	NLP	GenAI LLM	
hat	Appraisal	Digital curation	AI/ML GenAI Computer vision (CV)	

- **+150 papers since 2016**
  - IEEE Big Data Conference CAS workshops
  - ACM Journal on Computing and Cultural Heritage (JOCCH) Special Issue on CAS
  - Records Management Journal (Emerald Insight) Special Issue on Technology and Records Management
  - Miscellaneous journal and book chapters including Proceedings of Science, Archives and Primary Source Handbook, DigitalHeritage, UCL Press, Libraries, Archives, and the Digital Humanities, AI & Society Journal of Knowledge, Culture and Communication, and International Conference on Metadata and Semantics Research).
  -

Angles of Analysis

Articles tagged in the following key areas:

Archival Functions

- 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

Extended CAS Topics

- AI/ML
- Computer vision (CV)
- GenAI
- LLM

Record Types

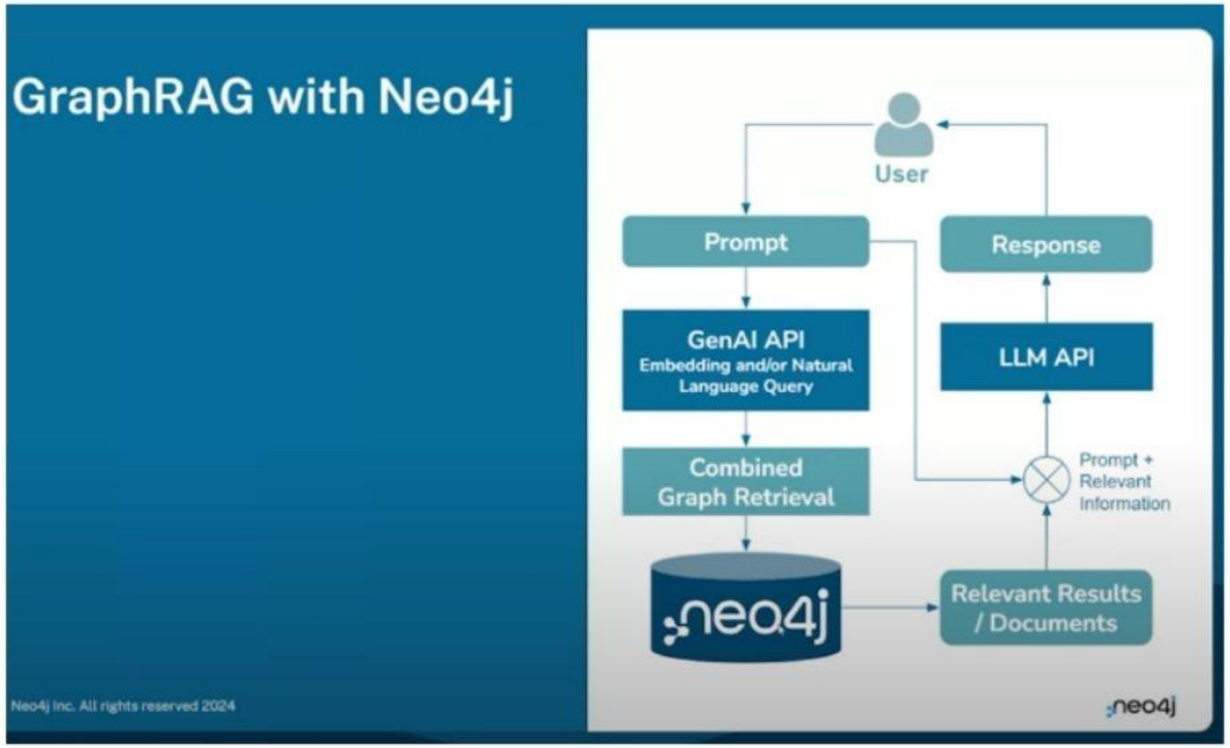
- Architectural & Engineering Drawings
- Artifacts
- Data Files
- Maps & Charts
- Moving Images
- Photographs & other Graphic Materials
- Sound Recordings
- Textual Records
- Web Pages

Social Justice Collections

CAS projects have engaged with a variety of fascinating collections including those highlighting:

African American voices
Black women's 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
WWII Japanese American Incarceration

GenAI, RAG, Knowledge Graphs



records.txt — Edited

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".

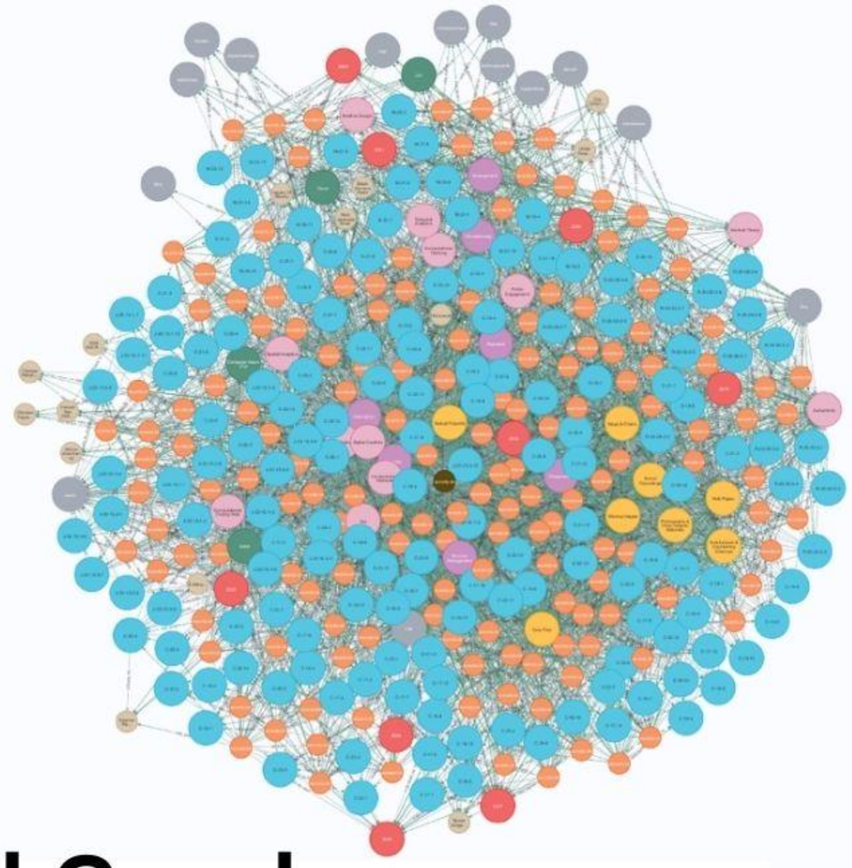
In the Year: "2018", Paper: "M-18-11" was Published in the Journal: "DigitalHeritage". It connects to the following Steps: "Description, Access", relates to the following Tools: "Iterative design, Digital curation, Computational methods", speaks to the following Topics: "WWII Japanese American Incarceration",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

In the Year: "2024", Paper: "M-24-12" was Published in the Journal: "ARCHIVES". 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".

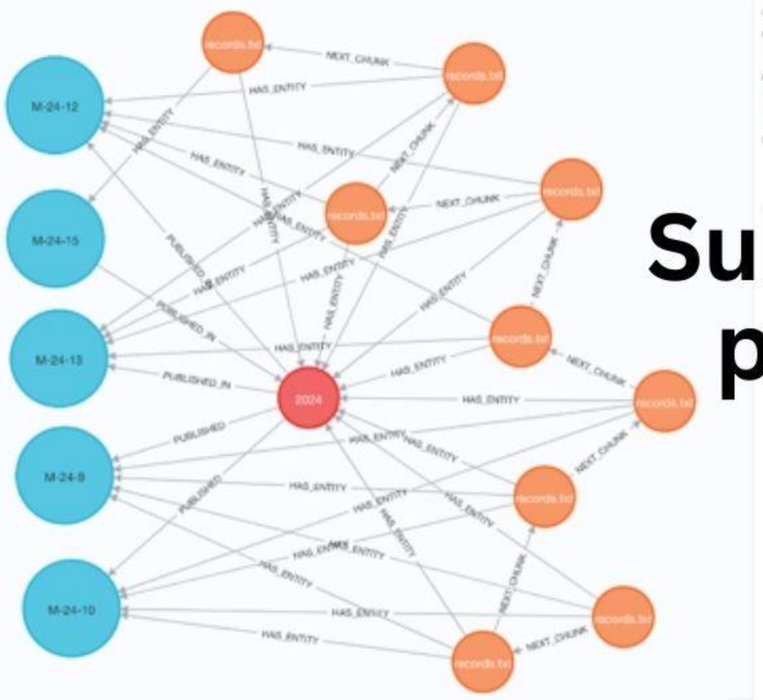
In the Year: "2024", Paper: "M-24-13" was Published in the Journal: "ISGC". It connects to the following Steps: "Accessioning, Arrangement, Description, Access", relates to the following Tools: "Iterative design, Digital curation, Spatial analytics, Temporal analytics", adds to the following Technologies: "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".

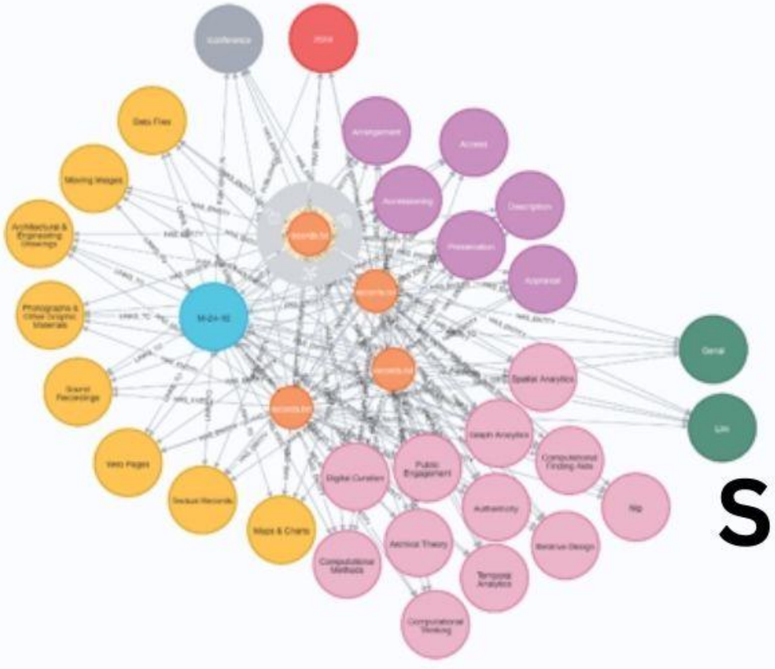
In the Year: "2024", Paper: "M-24-15" was Published in the Journal: "MTSR". It connects to the following Steps: "Description, Access", relates to the following Tools: "Spatial analytics, Digital curation", adds to the following Technologies: "AI/ML, Computer vision (CV)", speaks to the following Topics: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



Full Graph



Subset - 5 papers



Subset - Tags Used

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
Details

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
Details

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
Details

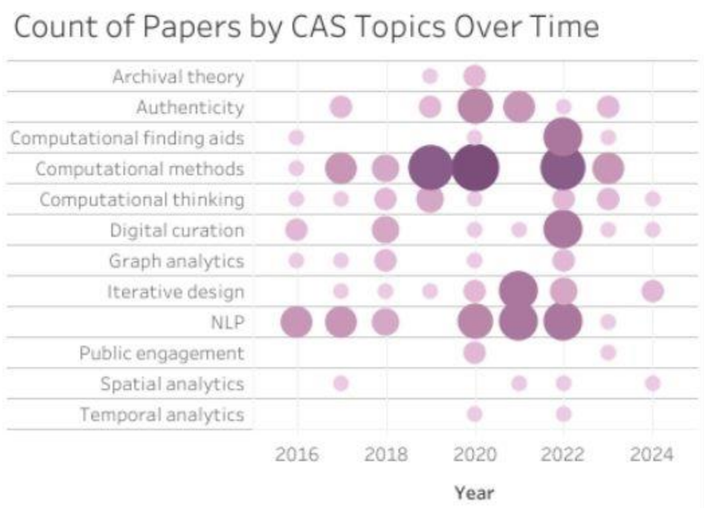
From these paper IDs which Tools are there?
10/14/2024 6:06:20 PM

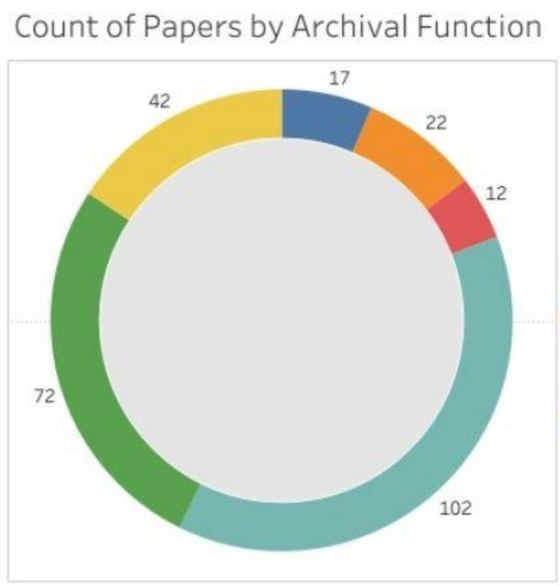
The paper IDs you mentioned are associated with the following Tools:
M-24-10: Computational Methods, Computational Thinking, Digital Curation, Graph Analytics
M-24-12: Digital Curation
M-24-13: Digital Curation
M-24-15: Digital Curation
M-24-9: Computational Thinking, Digital Curation, Graph Analytics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

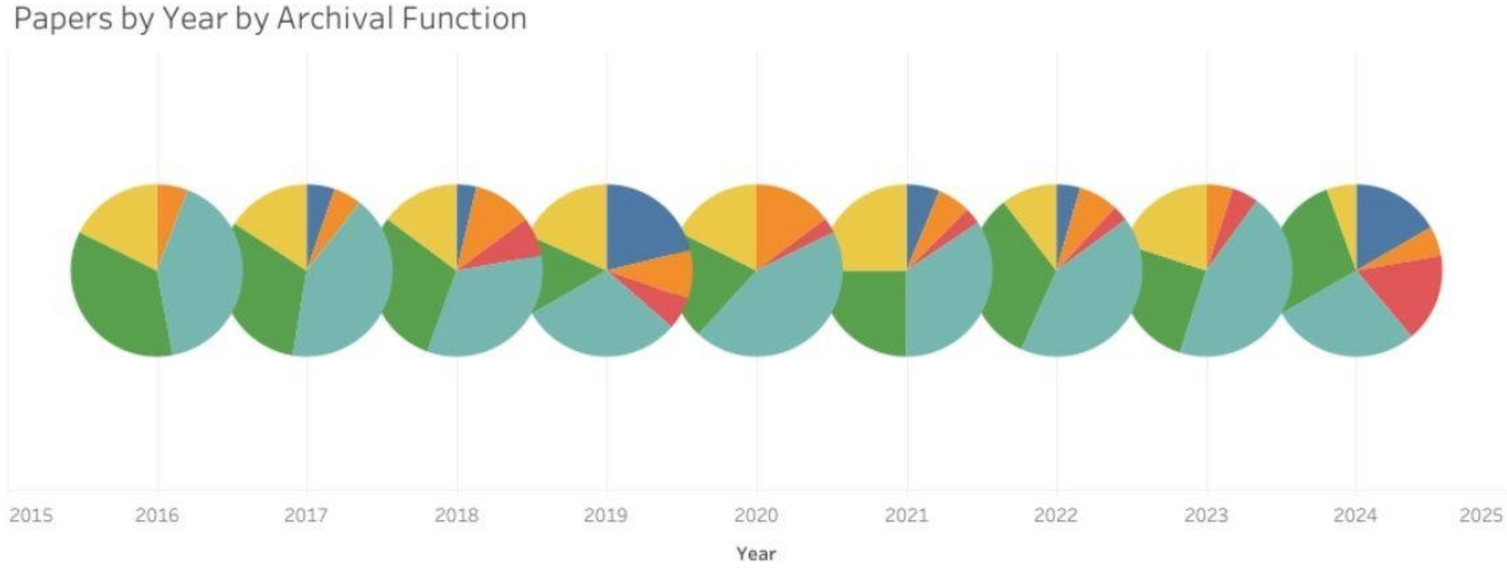
Interacting with the Data

Computatonal Review

Digital Analytics of collection data



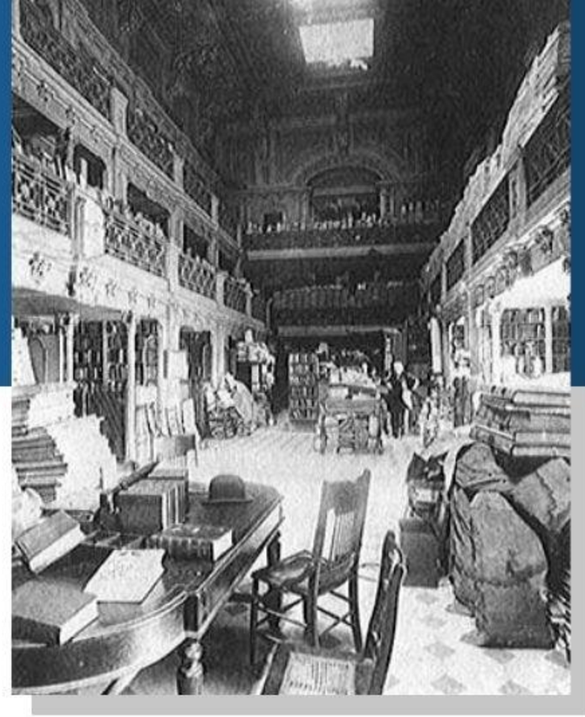




The Digital Tsunami

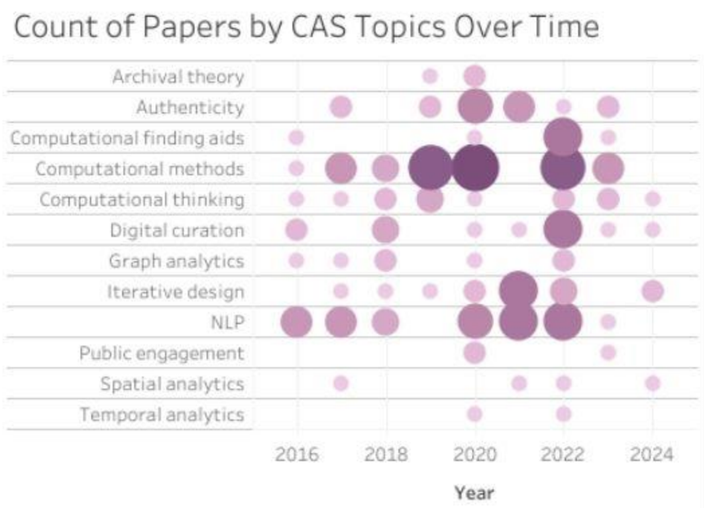
Backlogs are a consistent theme in Library and Archival Science. Pictured - the Library of Congress (stacks and unprocessed) in the 1890s. Digitizing both processed and unprocessed materials represents a significant challenge.

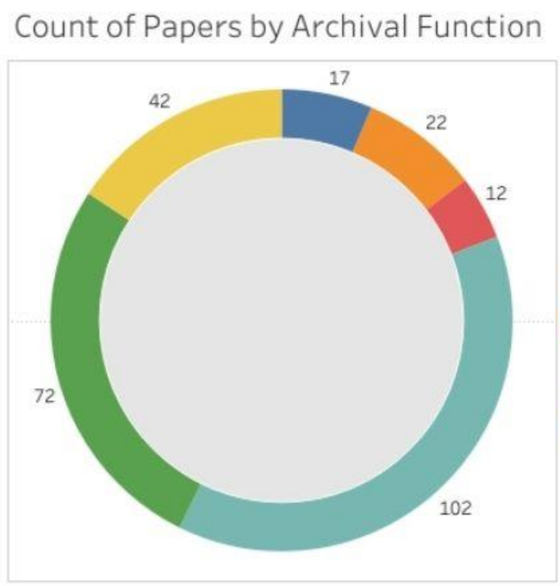
Computational Tools help us address challenges of (1) Digital Fragility, (2) Scale/Speed/Accuracy of Description, and (3) Connectivity / Engagement / Democratization of Access.

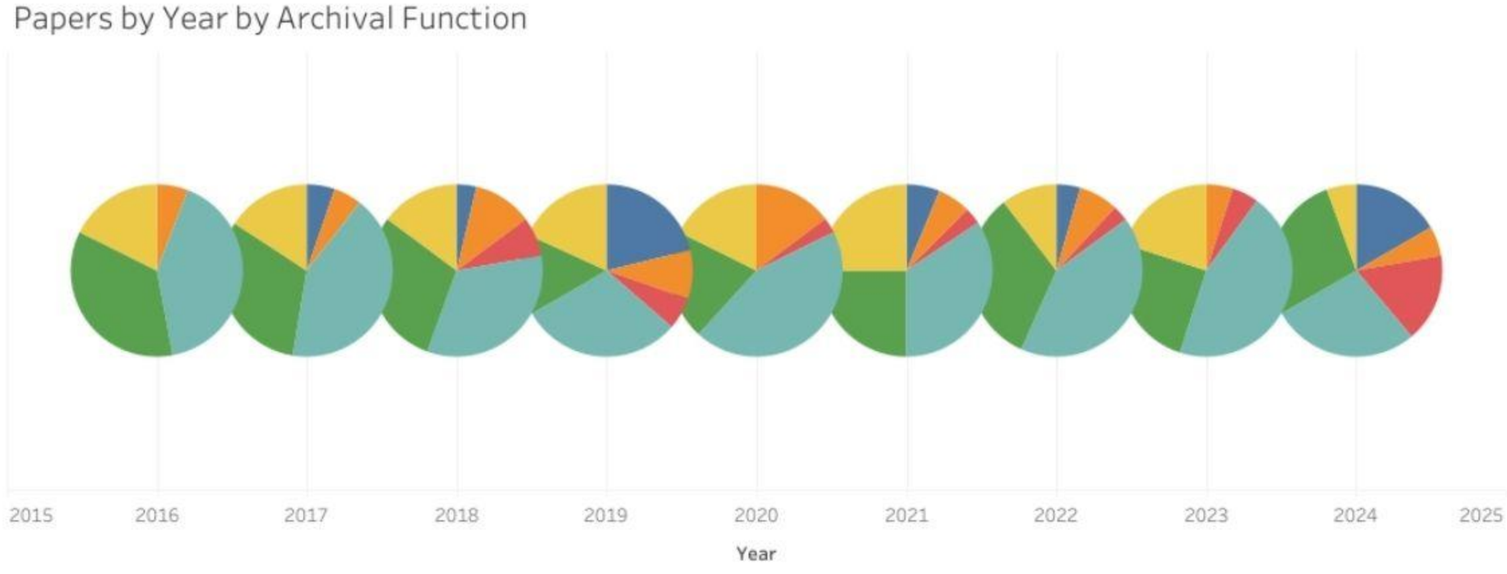


Computatonal Review

Digital Analytics of collection data

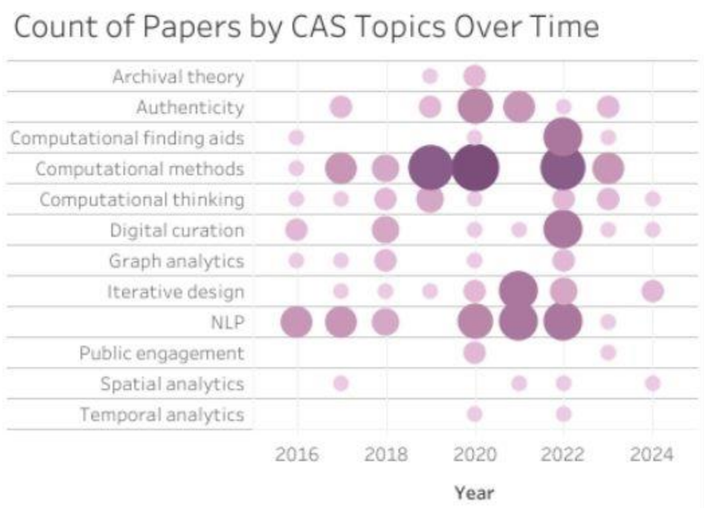


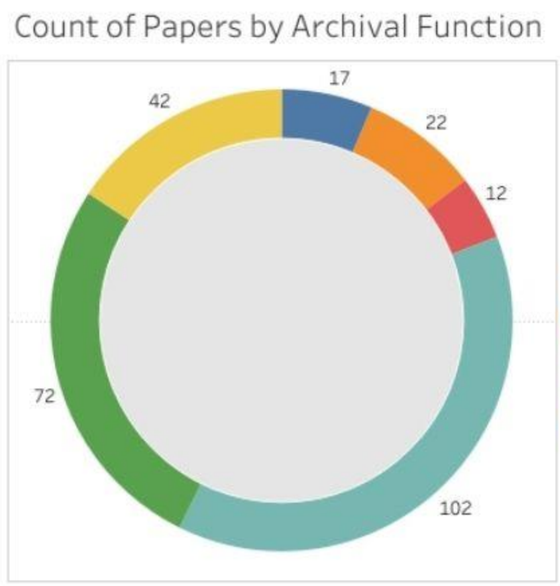


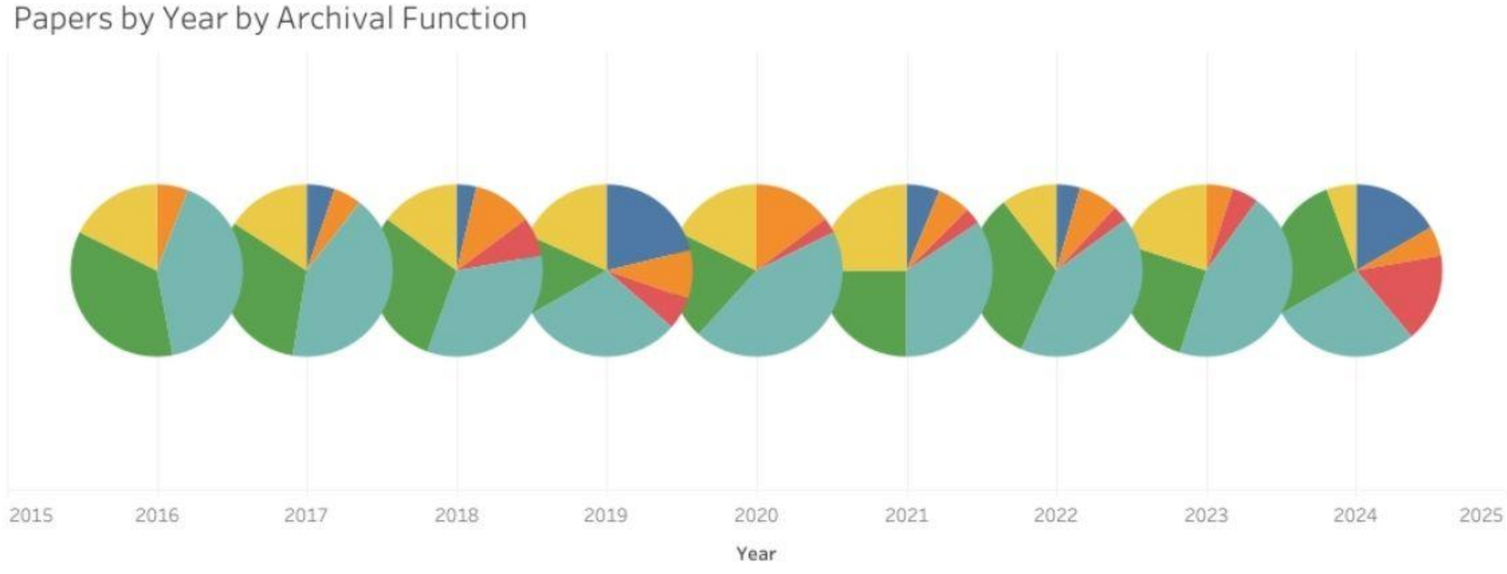


Computatonal Review

Digital Analytics of collection data

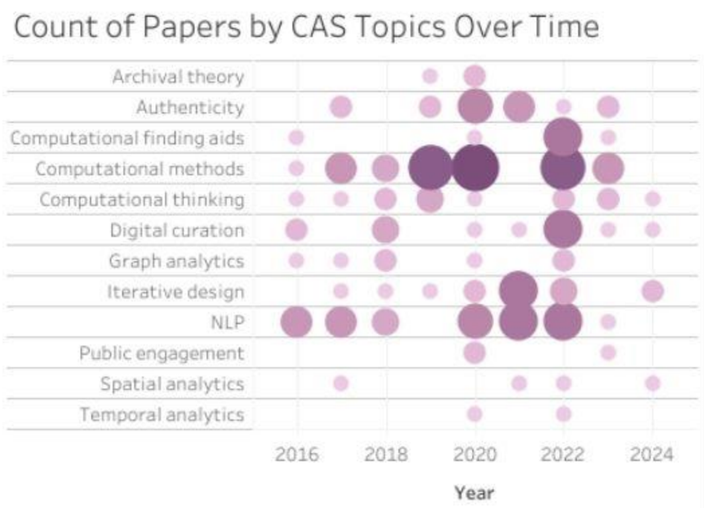


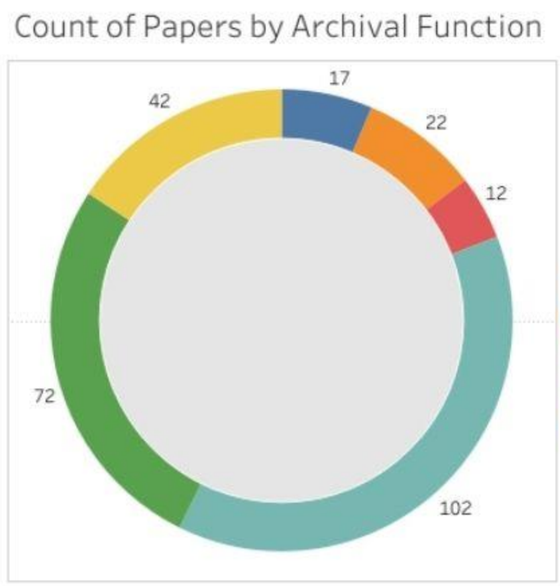


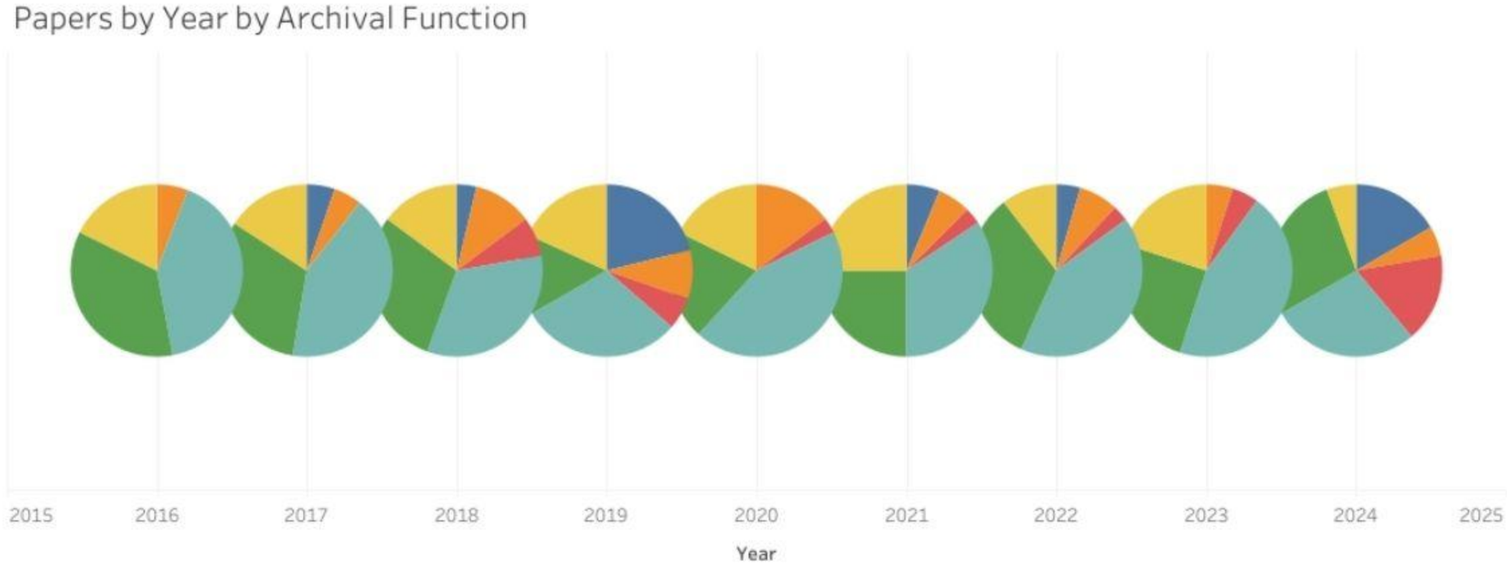


Computatonal Review

Digital Analytics of collection data



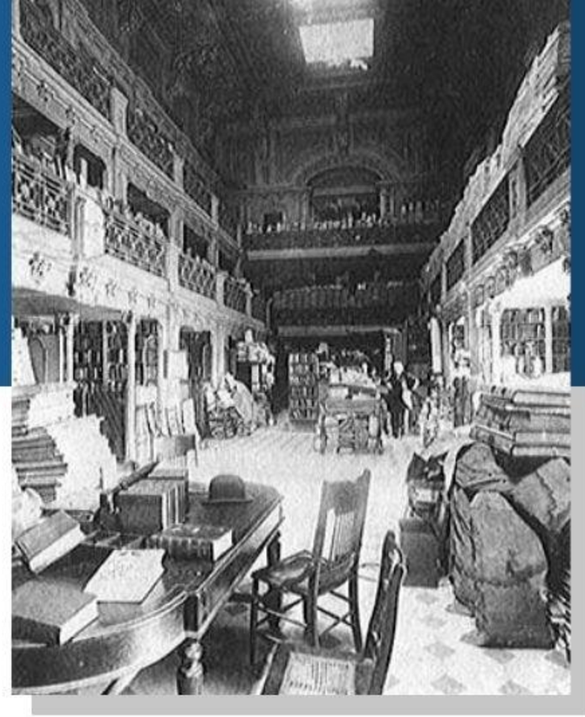




The Digital Tsunami

Backlogs are a consistent theme in Library and Archival Science. Pictured - the Library of Congress (stacks and unprocessed) in the 1890s. Digitizing both processed and unprocessed materials represents a significant challenge.

Computational Tools help us address challenges of (1) Digital Fragility, (2) Scale/Speed/Accuracy of Description, and (3) Connectivity / Engagement / Democratization of Access.



The Vanishing Box Problem

Born Digital Workplaces retain records very differently than paper ones.

- IT zips a desktop all together, without the Box level of structural order
- Our methods, as much as or more than our tools, must change, especially our advocacy

Computatonal Review

Digital Analytics of collection data

The Vanishing Box Problem

Born Digital Workplaces retain records very differently than paper ones.

- IT zips a desktop all together, without the Box level of structural order
- Our methods, as much as or more than our tools, must change, especially our advocacy

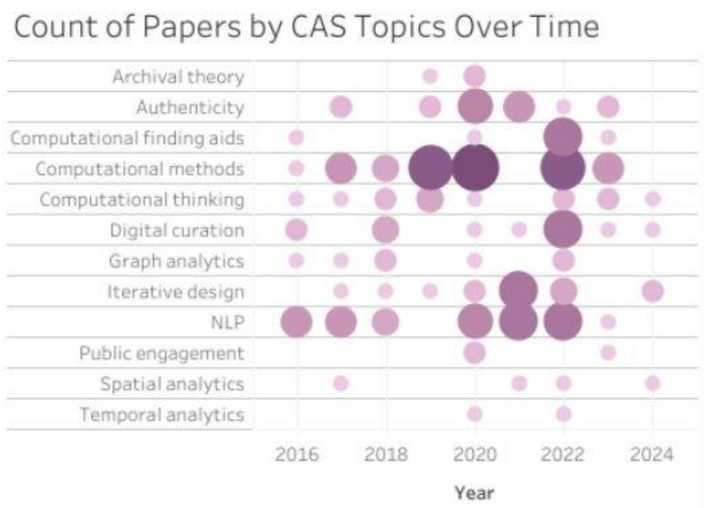
The Vanishing Box Problem

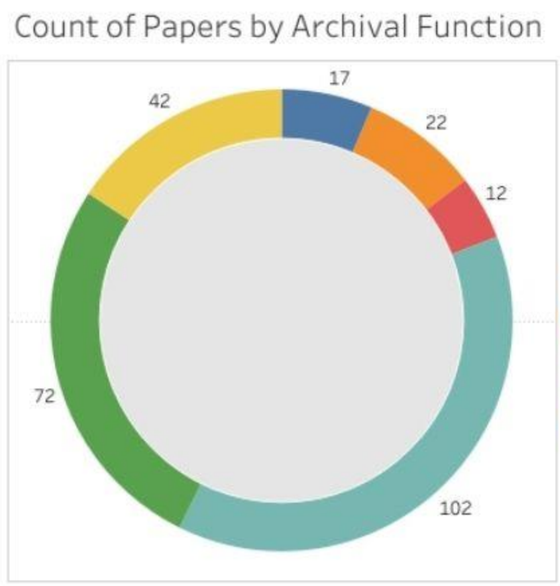
Born Digital Workplaces retain records very differently than paper ones.

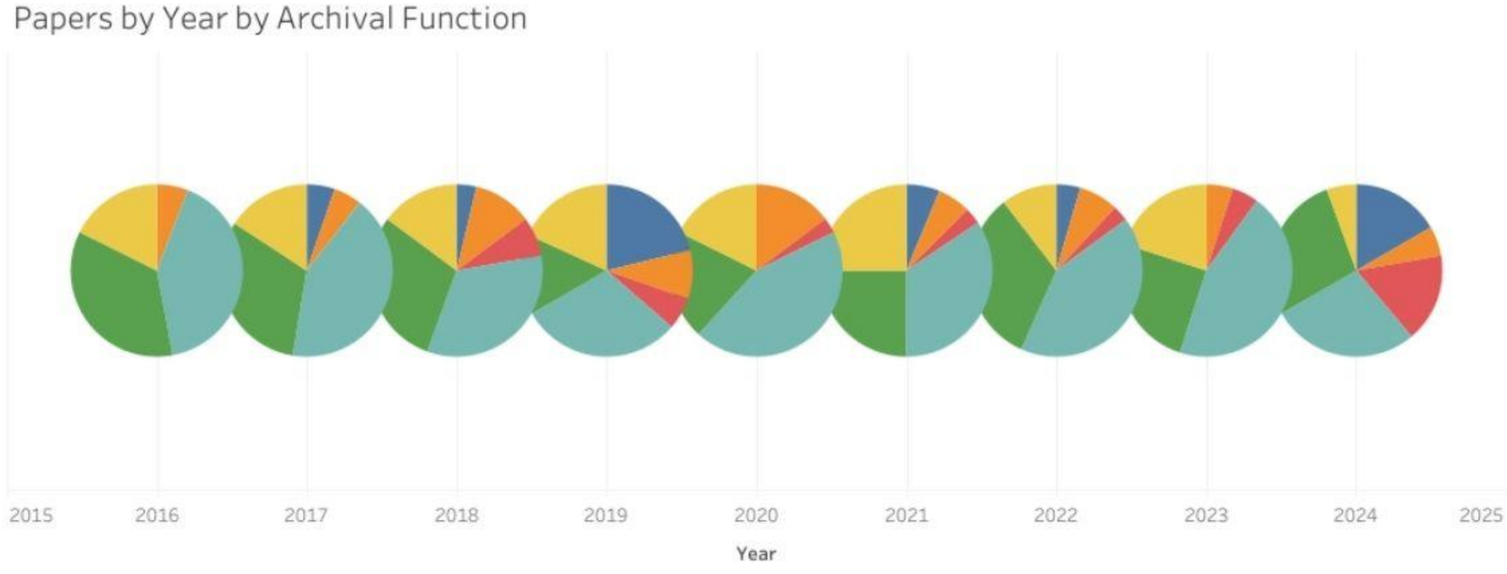
- IT zips a desktop all together, without the Box level of structural order
- Our methods, as much as or more than our tools, must change, especially our advocacy

Computatonal Review

Digital Analytics of collection data

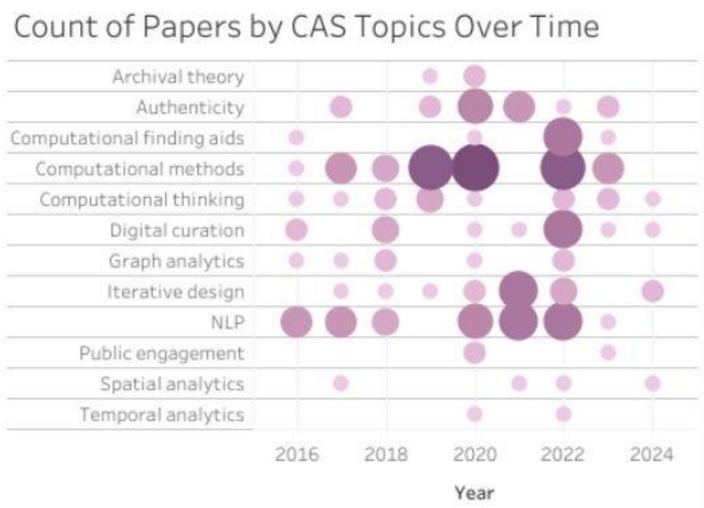


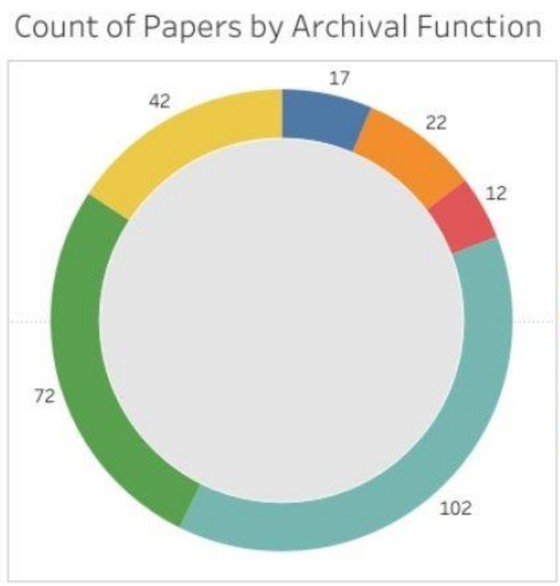


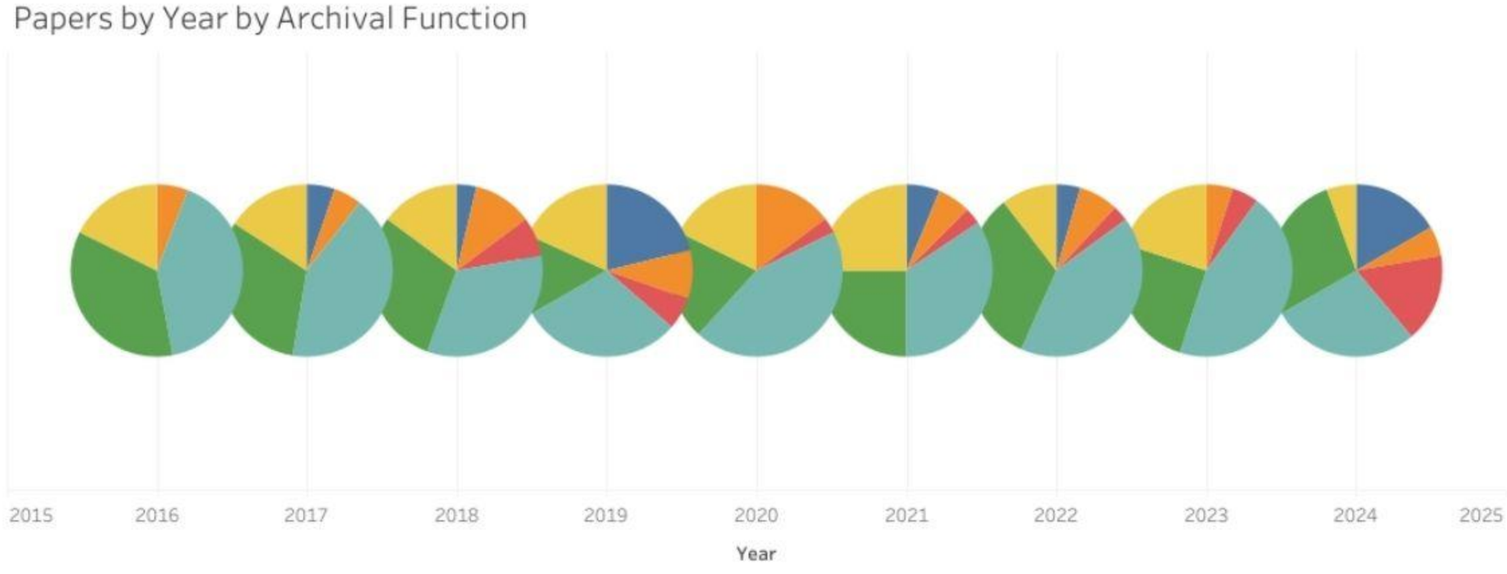


Computatonal Review

Digital Analytics of collection data







A Computational Review of the Literature of Computational Archival Science (CAS):

Advancing Archival Theory in the Age of the Digital Tsunami and the Vanishing Box Problem

Jennifer Proctor
Applied Research Laboratory for Intelligence and Security
University of Maryland
College Park, USA
jeproc@umd.edu

Richard Marciano
College of Information
University of Maryland
College Park, USA
marciano@umd.edu

https://ai-collaboratory.net/