Teaching Computational Archival Science: Context, Pedagogy, and Future Directions

Victoria L. Lemieux (University of British Columbia)
Richard Marciano (University of Maryland)

Abstract

The paper describes the development of the new transdisciplinary field of *Computational Archival Science* (CAS) and the significance of integrating computational thinking (CT) concepts into archival science. The authors illustrate the introduction of CAS into graduate archival training through two case studies at the University of Maryland and the University of British Columbia and discuss the implications of building and sustaining CAS educator networks.

The paper argues that, given the increasing use of AI in archival work and research, the acquisition of computational skills and competencies is urgent for those entering the profession, but sees several barriers, including the willingness of archival educators to engage in this space, the shortage of CAS educators, and the dearth of institutional support to help them acquire the necessary knowledge and skills. Additionally, there is a perceived conflict among some in the archival profession between CAS and recent archival scholarship emphasizing postcolonialism and progressive themes. The paper suggests that this is a false dichotomy, as demonstrated by the many CAS papers focusing on ethical and social justice aspects of the intersection of computing and archival work.

The paper concludes that the teaching of CAS is not merely an option but a necessity for the archival profession to stay relevant and responsive to the changing landscape and offers CAS graduate curriculum learning guidelines to address the need. and equip professionals with the tools and perspectives needed to navigate the complexities of digital recordkeeping, ensuring that archives remain accessible, trustworthy, and reflective of our evolving society.

Keywords: Computational Archival Science, Archival Training, Computational Thinking, Artificial Intelligence, Future of Archives

1. Introduction

Computing technology has been transforming the nature of archival work for decades. Initial transformations began with the rise of born-digital records in archives following the widespread adoption of computers and digital technology starting in the latter half of the 20th century and continuing apace even now. Additionally, the widespread adoption of computers and digital technology in the latter half of the 20th century opened the door for archivists to apply new computational practices and tools to archival work. One of the most significant changes has been the digitization of archival records. More recently, the application of artificial intelligence, for example, to the preservation, metadata generation, and identification of personal information has been transforming archival work (see, e.g., National Archives, 2016; Jaillant, 2022; Baron, Sayed & Oard, 2023). The rise of the internet, the web, and online databases have also made born-digital and digitized archival materials more accessible to a global audience. Researchers and the general public can now access archival collections from anywhere with an internet connection, reducing the need for in-person visits to physical archives. Additionally, advanced computing technology and techniques such as text mining, natural language processing, network analysis, and geospatial analysis have allowed researchers and archivists to perform data analysis and visualization on large sets of archival data to help uncover patterns, trends, and insights within collections, offering new perspectives and insights for researchers (Graham et al., 2016; Milligan, 2019). The many novel affordances brought about by the introduction of computing technology have also introduced novel challenges for archivists, however. Understanding the nature of new forms of born-digital records, such as those created using artificial intelligence or blockchain technology, remains an open challenge, as does how best to manage and preserve all types of born-digital records. In addition, the public and researchers are increasingly wishing to engage in new ways of researching archival materials that are potentially disruptive to traditional archival theories and practices (Marciano et al., 2019b). Archivists must now navigate new complexities related to the entrenched issues of digital privacy, copyright, and authenticity of records.

It is against this background of the transformative effects of computing technology on archival practice and theory, that the designation of *Computational Archival Science* (CAS) was devised in 2016 by the two authors of this paper with five additional co-authors (Mark Hedges from King's College London, William Underwood formerly from GeorgiaTech, Michael Kurtz formerly from the U.S. National Archives, Mark Conrad formerly from NARA, and Maria Esteva, from the U. Texas Austin). A Foundational paper was drafted and published in 2018 (Marciano, 2018a), making a case for a new transdiscipline for CAS. In 2018, Nathaniel Payne, at the time a doctoral student of Victoria Lemieux, further proposed a revised definition of CAS as follows (Payne, 2018):

> "A *transdisciplinary field grounded in archival, information, and computational science that is concerned with the application of computational methods and 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 and design, precision, and human technology interaction in support of acquisition, appraisal, arrangement and description, preservation, communication, transmission, analysis, and access decisions.*"

This chapter explores the kinds of CAS educational efforts and pedagogies that are emerging as part of archival science responding to the latest algorithmic and computational developments.

2. Context and Development of Computational Archival Science

The field of CAS has over 170 papers[1] to date. To promote the development of CAS, nine major CAS workshops were held as part of the international IEEE Big Data Conference with 119 papers presented, resulting in new CAS pedagogies (https://ai-collaboratory.net/cas/). CAS contributions alone represent five continents and 26 countries.

We exclude the growing number of papers that address the major themes and objectives of CAS that do not explicitly identify as CAS papers; these include books and papers generated by the AEOLIAN network[2] and the InterPARES AI project[3], to name only two examples. Combined with explicitly CAS papers, contributions from these projects send a strong signal that the archival field is taking a computational turn, albeit one with its own unique progressive and human-centered approach, as we will come to later.

One of the most important lessons learned has been the significance of integrating computational thinking (CT) concepts into archival science, which parallels the case for the inclusion of CT in mathematics and K-12 science classrooms. We will next describe a body of work on this topic that started in 2018. The emergence of Computational Archival Science parallels that of Computational Social Science, Computational Biology, and more recently Computational Journalism, with a specific focus on archival processing and theory.

(Underwood et al., 2018). (Marciano et al., 2018b) are the first papers where working with LIS students, we introduced a CT framework organized as a taxonomy of 22 CT practices organized around four categories. We showed how to map these practices through an archival case study involving World War II Japanese American Incarceration Camps focusing on automating the detection of personally identifiable information or PII.

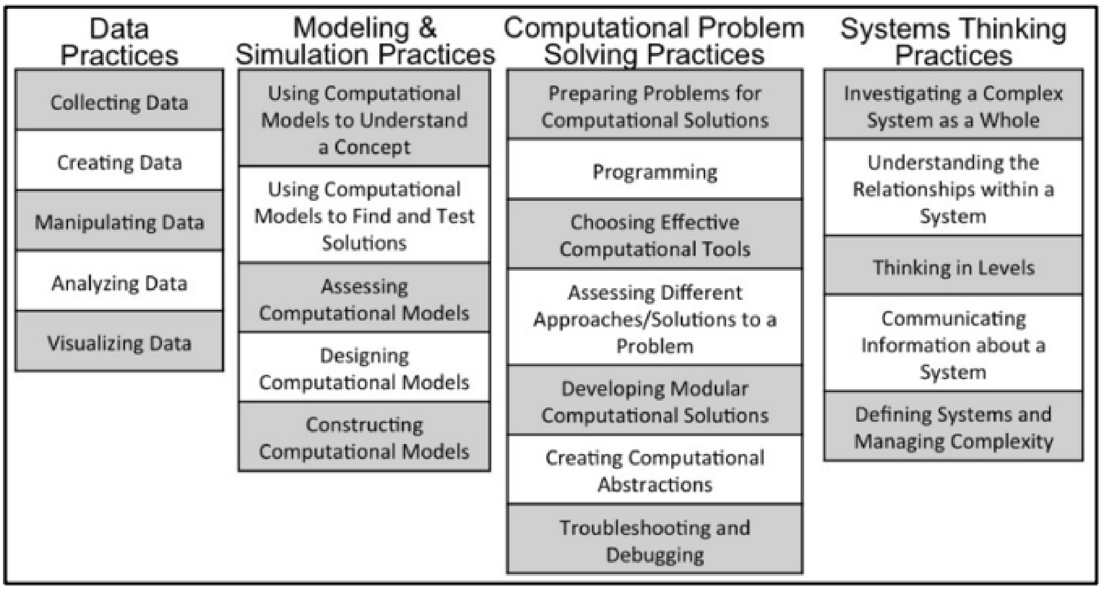

Figure 1: Computational thinking practices taxonomy.

A subsequent larger study (Marciano et al., 2019a), involving 20 students (undergrad and grad), mapped five case studies to the CT taxonomy: (1) Detecting personally identifiable information, (2)

[1] Compendium of Core Computational Archival Science (CAS) Papers: 4 tabs, https://docs.google.com/spreadsheets/d/1oCVCWpik_zjdilih9iXh2KITZ7LIOVIt/edit?usp=sharing&ouid=105652788602997060030&rtpof=true&sd=true

[2] AEOLIAN network: https://www.aeolian-network.net/.

[3] InterPARES Trust AI: https://interparestrustai.org/

Developing name registries, (3) Integrating vital records, (4) Designing controlled vocabularies, (5) Mapping events and people, and (6) Connecting events and people through networks.

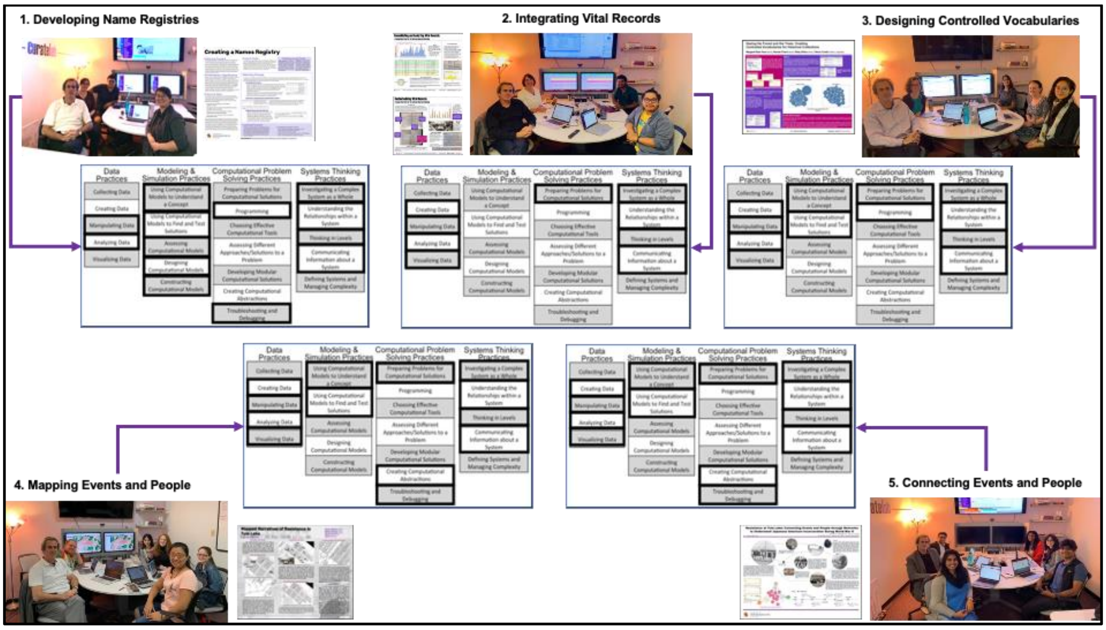

Figure 2: The CT taxonomy in practice with MLIS students.

The value of mapping diverse CAS case studies into a common CT framework is clear as it provides unity and coherency in identifying the underlying computational skills and techniques that are relevant to archival work and a common language for teaching CAS to students.

Furthermore, what we were able to establish is that all 22 CT practices can be mapped into archival science case studies (Underwood et al., 2019a; Underwood, 2019b), demonstrating the generalizability of this approach. It is possible that in the future, in mapping in the other direction (i.e., from new computational archival projects to the 22 CT practices), some case studies may challenge the completeness of this original taxonomy. Furthermore, Buchanan et al., 2022 demonstrate experimentation with CT in archival science courses across four campuses and over 3 semesters of networked instruction.

As we write, the focus, goals, and disciplinary approaches within CAS remain unsettled and continue to evolve. Ongoing research into, analysis, and discussion of the focus, goals, and techniques of CAS as an emerging discipline will continue to play an important role in defining the competencies appropriate for students of CAS and the pedagogical content and techniques that best deliver those competencies, which we discuss in the next section.

3. Introducing CAS into Graduate Archival Training: Case Studies

a. University of Maryland

In 2015, one year before the start of the CAS initiative, Richard Marciano and Michael Kurtz launched two pre-CAS pedagogical initiatives at the U. Maryland, the Digital Curation Innovation Center (DCIC) (link) and the Digital Curation for Information Professionals (DCIP) (link).

The DCIC aimed to create a hands-on lab culture for LIS students, promoting interdisciplinary projects that explored the integration of archival research data and technology to generate new forms of analysis for historical research engagement. Projects were structured around social justice and infrastructure themes (link). Student-led "Datathons" included an 8-week UMD Digital Curation sprint on Japanese-American WWII Camps (link) and a 2-day Maryland State Archives (MSA) datathon on Legacy of Slavery (link). The DCIP was the U. Maryland iSchool's first professional certificate program designed for individuals currently in the library, archives, or digital curation field, or who were planning to enter it but wish to enhance their digital skills. The program (2015-2024), through three online courses over 9 months, took students from introductory through advanced digital curation lessons while focusing on practical applications and a final hands-on capstone project. Examples of projects include 2021 (link), 2022 (link), and 2023 (link), and 2024 (link). In 2020, with the COVID pandemic, Marciano and Lemieux joined forces with colleagues and pivoted to an online virtual research network, the Advanced Information Collaboratory (AIC) (link). The focus is on: (1) exploring the opportunities and challenges of "disruptive technologies" for archives and records management (including digital curation, ML, AI), (2) leveraging the latest technologies to unlock the hidden information in massive stores of records, (3) pursuing multidisciplinary collaborations to share relevant knowledge across domains, (4) training current and future generations of information professionals to think computationally and rapidly adapt new technologies to meet their increasingly large and complex workloads, and (5) promoting ethical information access and use.

Computational activities are now being incorporated into regularly scheduled LIS graduate courses. Our goal is to modernize archival and library education and contribute to the development of faculty and library "digital leaders." Examples include:

- INST747: **Research in Advanced Digital Curation** (see Jun. 18, 2022 student project link)
- INST604: **Introduction to Archives and Digital Curation** (see Dec. 5, 2023 final project link)
- INST742: **Implementing Digital Curation** (see May 6, 2021 datathon link, May 5, 2022 showcase link, May 11, 2023 datathon link ,and May 9, 2024 datathon link).
- **Spatial, Graph, GenAI & LLM Analysis**: In the summer of 2024, we introduced new MLIS courses on these topics applied to archives (link).

Figure 3: A Knowledge Graph for urban renewal in Asheville, NC (*Author: Nick de Raet*)

Computational exposure was also incorporated into undergraduate Information Science courses. Examples include:

- INST341: **Introduction to Digital Curation** (see Dec. 9, 2021 link).

Incorporating computational research and training has led to media outcomes for students and portfolio projects they can incorporate into their CV, such as the followings:

- **Terp Magazine** Winter 2023 – "Truth in Exile" (link), with a project featured on **NHK World Japan television** (link)
- Public access portal that documents the hidden legacy of urban renewal in Asheville, NC (link), and March 20, 2023, public testimony to the **Asheville Community Reparations Commission**, Mayor, and City Council (link).

Increasingly, we are exposing students (LIS grad and undergrad) to data science and archives through the creation of Jupyter Notebooks. See student feedback (Piety, 2023), with presentations and notebooks (link). The ultimate goal is to contribute to the development of faculty and library digital leaders. In addition to preparing our students to meet these challenges through research-oriented seminars, we are building an online repository called CASES (Computational Archival Research Educational System) involving CAS practices that address practical digital records management and archival problems (https://cases.umd.edu/). We are exploring how computational thinking practices might be introduced to graduate students in the core Archival Studies curriculum. The goal is to enable a collaborative network of educators and practitioners who can learn from one another through the sharing and dissemination of computational case studies and lesson plans.

b. University of British Columbia

Teaching in CAS began at the University of British Columbia (UBC) in a course developed by Victoria Lemieux in 2011 on Information Visualization and Visual Analytics. The course, which began as a highly experimental topics course in its early days and continues to this day, now has a permanent place in UBC's Masters of Archival Studies (MAS) and Master of Library and Information Science (MLIS) programs. Since the takeaways from this early CAS work for archival education have already been discussed (Marciano et al., 2018a), here it is sufficient to remark that the content of this course was one of the first archival courses to expose students to the application of machine learning, artificial intelligence, and data visualization to archival work and research, Through the subsequent writing of Devon Mordell, who was a student in one of the offerings of the course, it also was able to seed the formulation of a new "archives-as-data" paradigm (Mordell, 2019).

Experimentation with CAS curriculum and pedagogy continued at UBC with the introduction of a summer institute on blockchain technology in 2017 and the formal introduction of a Blockchain Graduate Pathway in 2019. The program has been recognized by Coindesk as number one in Canada for blockchain education (Coindesk, 2023). In its current formulation, the blockchain graduate pathway aims to provide students with the knowledge and skills to contribute to the advancement of research and development in blockchain technology while at the same time engaging in some of the world's most complex sociotechnical issues. The multidisciplinary Blockchain Graduate Pathway comprises a 12-credit non-degree training program that augments existing master's and PhD programs at UBC. Students in the Pathway take courses on "Issues and Perspectives on Blockchain Technology"; technical "Foundations in Blockchain and Distributed Ledger Technology"; Blockchain and Distributed Ledger Technology research; and "Blockchain for Information Professionals," which delves into various records and information-related aspects of blockchain technology, such as their nature as records and record keeping systems, archival theoretic critiques of blockchain solution design, and compliance with regulations, including privacy, data protection and copyright.

In addition to completing coursework, students complete a research-based industry internship and prepare a major research paper or thesis based on the research conducted for the internship. The following papers represent examples of outcomes of this work: (Lemieux, Hofman, Batista, Joo 2019; Kang 2021; Lemieux, Voskobojnikov & Kang 2021; Suleman 2022; Suleman & Lemieux, 2023).

MAS and MLIS students who have completed this program have gained knowledge and experience equipping them to fill a range of archival, records management, and data and information management roles, as attested to by this former student:

> "*I first heard of the Blockchain @ UBC program in the iSchool newsletter. Having had no prior understanding of Blockchain beyond knowing it had something to do with Bitcoin, I was interested in knowing more. The program provided me with an interdisciplinary knowledge of Blockchain technology and the countless industries and use cases it can be applied to. Through the course work, the Summer Institute, and an internship with the Land Title and Survey Authority of British Columbia, I was able to apply theoretical concepts in the real world. While I currently do not work within the Blockchain space, the transferable skills I gained through the program allowed me to become an information specialist at a technology company a few months prior to graduation.*[4]"

The final example in this case study of integrating CAS into the curriculum for students of archives is a recently redesigned offering of a first-year core graduate course on archives on technology (ARST 500). The course aims to provide knowledge of the role of technology in archival work, at the theoretical and pragmatic level. Specific learning outcomes of this course are as follows:

- Discuss the information and communication technologies in use in archives work and the affordances, both positive and negative, of these technologies
- Critically evaluate the differences between archival and computational theories and methodologies
- Apply knowledge of technology in support of archival work (e.g., set up a database and a website, use natural language processing, machine learning, and data visualization to analyze and make available information about archival materials
- Critically reflect upon the application of computational techniques and tools to archival documents or for archival work

Themes of archives as data; comparative notions of integrity and authenticity; knowledge representation; power and positionality; and privacy are also explored in this course.

One of the primary methods by which the learning outcomes are achieved is by assigning students a series of labs designed to help them explore the application of technology in support of archival work, and the implications of doing so in concrete and practical terms. In the 2023 offering of the course, groups of students were asked to conduct Natural Language Processing and Sentiment Analysis on supplied sample text extracted from a 1920 newspaper. To complete the lab, they used Google Colabs, a free Jupyter Notebook environment that requires no setup and runs entirely in the cloud, which eliminated the need for students to download software or set up a virtual environment to run the lab. Google Colab was used to run the Transformers Library by Hugging Face using the DistilBERT model. Transformers offers an open-source library that provides state-of-the-art machine learning models, primarily for tasks in Natural Language Processing (NLP). As students were not expected to have prior programming experience, and the pedagogical purpose of the lab was not to teach them programming but rather to expose them to processes of datafication of archives, how natural language processing works, and the implications of using such

[4] UBC iSchool, Blockchain Graduate Pathway: https://ischool.ubc.ca/graduate/areas-of-focus/blockchainubc/

tools in relation to data/records integrity and bias, they were given a "recipe" containing the script in Python to run the sentiment classifier rather than being asked to write the code themselves.

To stimulate students' critical thinking on the application of sentiment analysis to archival material, the lab was paired with readings, including (Mordel, 2019) and (Jo & Gebru, 2020), which students were asked to connect back with their experience of conducting the lab.

Several groups went beyond the specific assigned task to conduct further experiments using different sentiment classifiers and text. For example, one student group ran an experiment using the SpaCy sentiment classifier on Chinese text, generally observing that the results were consistent with what the students would have expected the classification of the words to be. Another group ran the DistilBERT model over specific words relating to sexuality (see Figures 4), concluding that the results showed a "homophobic" bias and that the classifier was unable "to understand the multiplicity of meanings [of] a word…which is quite problematic." This example of experiencing various flavors of sentiment analysis serves the purpose of developing critical thinking skills that show the strengths and weaknesses of computational techniques.

	Text	Sentiment	Score
0	gay	NEGATIVE	0.995823
1	queer	NEGATIVE	0.990332
2	lesbian	NEGATIVE	0.989495
3	homosexual	NEGATIVE	0.987549
4	heterosexual	POSITIVE	0.924257

Figure 4: Sentiment Analysis for Sexuality applied to single-word tokens.

c. Building and Sustaining Educator Networks

Successful examples of emerging educator networks can be found in Buchanan et al. 2023), where LIS students developed archival case studies with Jupyter Notebooks that document the International Research Portal for Records Related to Nazi-Era Cultural Property (IRP2), and a selection of unclassified catalog entries about digitized nuclear science reports. Archival data visualizations are shown to promote outreach and access. Other examples include the LEADING Network (LIS Education And Data Science Integrated Network Group) focused on preparing a cohort of LIS doctoral students and early to mid-career librarians for data science endeavors (Greenberg et al., 2023).

Beyond these emerging educator networks, to support the education of archival practitioners who are competent in CAS, it is necessary for there to be willing learners and archival educators to teach. This has been a challenge, as noted in a recent presentation made by Anne Gilliland on the theme of "Preparing Archivists in Computational Thinking and Innovative Technologies" at the International Academic Conference on Digital Intelligence: Empowering the Modernization of

Archival Work, hosted by the School of History and Culture, Shandong University, China in October of 2023,

> *"It has been quite difficult to even develop the teachers of archivists, the professors who teach them, to develop those skills or even the awareness of where computational processes would be appropriate to use in archival work, archival work, as opposed to digital humanities, work or STEM. and I think it has been equally difficult to persuade students that this is not just an important part of archival work but an essential part of archival work."*

Our observation is that archival educators, by and large, lack the technical knowledge and skills to effectively teach CAS. Moreover, there is seldom institutional support to help them acquire the necessary knowledge and skills, as discussed in a recent paper by Lise Jaillant (2024). It takes incredible dedication and effort for someone "classically trained" in archival studies to acquire the technical capability to teach CAS. Equally, it takes training for anyone "classically trained" in technical disciplines, such as computer science or engineering, to acquire archival knowledge and skills to teach CAS. The CAS educational challenge lies not only in the absence of technical knowledge and skills, however, but also in a lack of technical curricular content and pedagogical knowledge of how to teach such material to learners with very different levels of technical knowledge, skill and aptitudes.

Perhaps a greater challenge is motivation, though: while the transformation in archival work wrought by computation should be motivation enough for archival educators to "grasp the nettle", as Gilliland observes, this does not always happen. Similarly, more technical disciplines may lack interest in working with cultural materials and fail to appreciate the many other applications of archival theory (e.g., identifying deep fakes, to name but one "au courant" area).

To address the shortage of CAS educators and the many challenges of providing CAS education, and to stimulate a meaningful dialogue among educators of diverse disciplinary backgrounds with a view to developing CAS as a novel transdiscipline, we are now engaging in building a TALENT Network for the Training of Archival & Library Educators with iNnovative Technologies). See: *https://ai-collaboratory.net/projects/talent-network/*. TALENT brings together experts from across the United States (including archivists, librarians, Library and Information Science educators, historians, learning scientists, cognitive scientists, computer scientists, and software engineers) to create a durable, diverse, and multidisciplinary national community focused on developing digital expertise and leadership skills among archival and library educators. The network focuses on four major objectives: (1) doubling the educator network to include diverse educators in multidisciplinary iSchools that have a focus on adjacent disciplines (e.g., computing, engineering, education, and data science) where there are substantial needs for archivists and librarians with digital and computational skills, (2) engaging HBCU students (Historically Black Colleges and Universities), (3) conducting curriculum development through experts from a Learning Sciences Network, and (4) addressing the social and ethical concerns that arise from computational and algorithmic thinking.

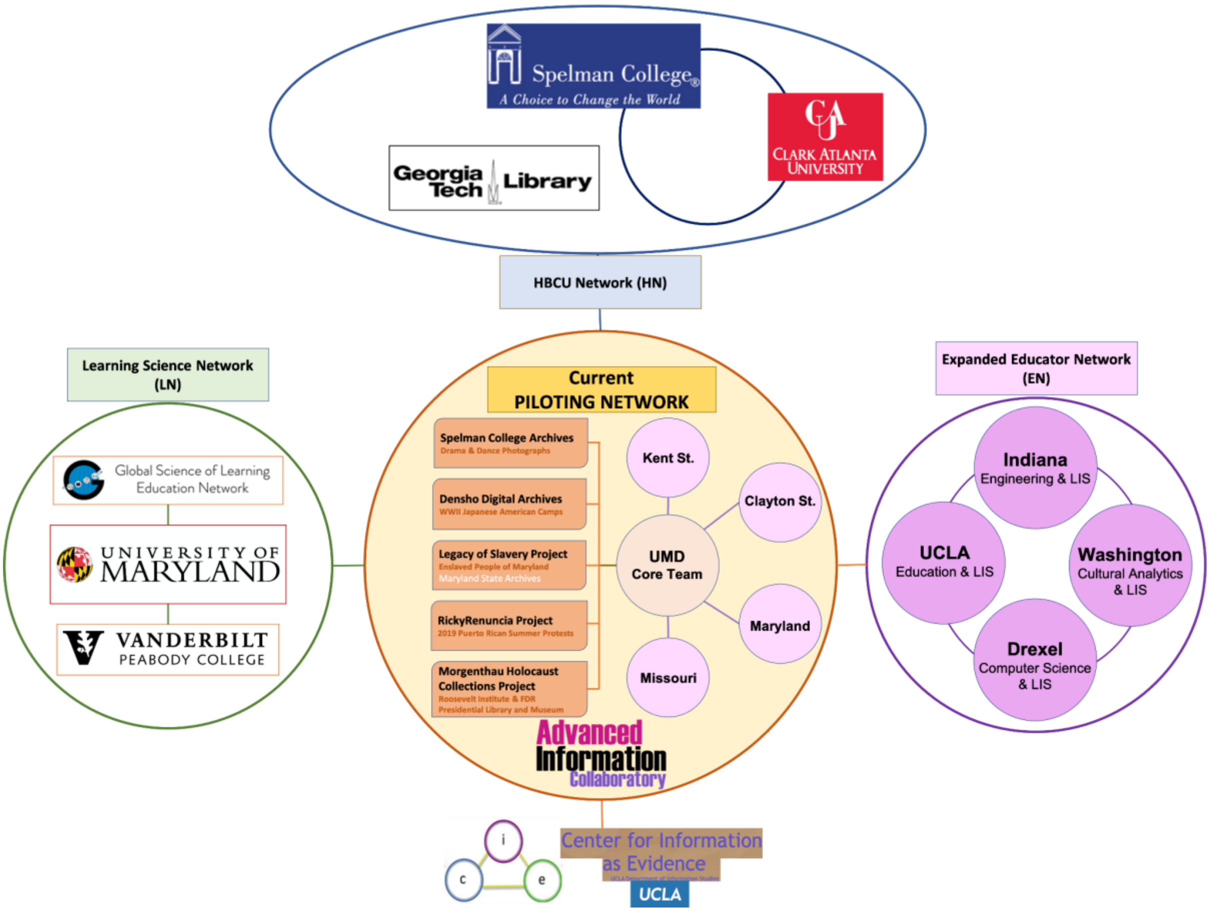

Figure 5: The TALENT Network to broaden the teaching of CAS.

All of the efforts to build and sustain CAS educator networks are for naught, of course, unless, as already mentioned, there are willing students. We note that in recent archival discourse, technology has often been associated with colonialism and other negatively perceived "isms", as this statement exemplifies:

> *In 2011, the Pluralizing the Archival Curriculum Group* (PACG) *of the Archival Education and Research Institute noted that "Archival studies education programs are conceptualized in strikingly similar ways worldwide, largely because of the overarching bureaucratically- and legally-centered paradigms developed and exported from Europe through colonialism, evangelism, mercantilism, and technological developments, and later codified through national and international standards and terminologies* (PACG, 2011).

Jaillant (2024) makes a similar observation about negative perceptions of technology. While efforts in pluralizing the archival curriculum are indisputably enriching, it is worth considering the potentially "chilling" effect such perspectives may have on students' willingness to learn about CAS and develop the technical skills that, as Gilliland (2023) and Jaillant (2024) observe, are not merely important but, nowadays, an essential part of archival work. This begs the question of how best to bring pluralizing perspectives into dialogue with CAS perspectives. Certainly, educational offerings focusing on the ethical aspects of applying technology to archival work, as discussed above, and those that exemplify how CAS techniques can be applied to advance social justice themes, help to achieve this end, but more effort is needed. The desire to advance this effort further is what, in part, motivated us to write this chapter, and also partly shaped our thinking on future directions for CAS, as we discuss in the following section.

4. Future Directions

In articulating a new "archives-as-data" paradigm, Devon Mordell (2018), points to some avenues for expansion of the CAS construct. She refers, for example, to the "invisibilization" of archives that can hide "the human decision-making within the often-opaque interfaces of computational tools." Her reflection suggests an opportunity in CAS to methodically dissect and understand the intricate, interwoven layers of technical frameworks. This approach serves two purposes: firstly, to aid archival studies students in comprehending both the theoretical and tangible components of computing infrastructures, and secondly, to critically examine the discursive aspects of data as well as the inherent power dynamics and cultural positionality manifest within computational infrastructure and the logics of computing. By preparing students to think critically about these aspects, the approach helps to create a generation of archival professionals who are more likely to avoid falling into the trap of believing that technology is the solution to meeting every archival challenge and who can also anticipate, identify, and address ethical challenges and societal concerns in the application of technology in their work.

Fearing "reactionary tendencies" in CAS, Mordell further counsels the archival profession to ensure that a social justice critique is maintained within CAS. Madelynn Dickerson and Audra Eagle Yun's 2023 CAS Workshop paper on "Critical Community-Centeredness: Ethical Considerations for Computational Archival Studies" signals how CAS might move in the direction of Mordell's recommendation. At the same time, the many papers highlighting the application of technology to advance social justice goals, such as Richard Marciano's work on the "Legacy of Slavery" and "Racial Reparations for Urban Renewal" illustrate how CAS techniques can uncover important insights to support processes that identify and help deal with the injustices of the past.

To further illustrate the impact of CAS on supporting past injustices, the Asheville Community Reparations Commission met on June 17, 2024, voting in support of a 10-page reparations request to the city amounting to $148K to families harmed by the displacement caused by urban renewal. A preliminary list of the names of individuals & businesses in the Southside neighborhood impacted and eligible for cash payments was compiled by students. See: https://docs.google.com/spreadsheets/d/110xsc3XXbm9Wa7klEJ75vnKw_b_BdimY/edit?gid=1080328214#gid=1080328214. This data was compiled using computational approaches as demonstrated in the resources listed in section 3.a. (public access portal, and public testimony links).

There also is room to use novel technologies to support archival educators in teaching CAS techniques, given the challenges of acquiring CAS knowledge and skills. In a recent CAS study, GPT-4's knowledge in some areas of archival practice, and its ability to think computationally about archival tasks are investigated (Underwood & Gage, 2023, Underwood & Gage, 2024). It is demonstrated that GPT-4 shows an understanding of 17 of the 22 CT practices. "These results support the possibility of ChatGPT+ being a partner to educators and students of archival studies in learning computational methods and tools applied to digital archival practice is a possibility. Similarly, archivists, records managers, and digital curators might be better able to apply computational methods and tools to their tasks with the aid of a ChatGPT+ agent." Ethical and trustworthy AI considerations should of course be part of this exploration, and indeed are a necessary component in developing computational archival pedagogies.

Finally, we think the time has come to move towards guidelines for graduate-level education in CAS. These guidelines should lay out the curricular content, pedagogy, learning outcomes, and graduate competencies to be developed. As a first effort, we propose the following set of competencies divided into three main areas of focus: technical, records/data, and social, based on the development of CAS to date and following Lemieux and Feng's (2021) "Three Layer Model" of system design:

CAS Graduate Competencies:

Technical

1. Graduates possess technical knowledge and understanding enabling them to apply computational thinking, tools and methodologies to effectively analyze and preserve digital records and support archival work. Specifically, graduates have the ability to:

   1.1 apply knowledge of technologies to real world record creation and recordkeeping problems and situations

   1.2 demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data

   1.3 demonstrate an understanding and skill in how to apply modeling and simulation practices to archival work, such as using modeling to understand a concept, using computational models to find and test solutions, assessing computational models, designing computational models, and constructing computational models.

   1.4 within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging.

   1.5 demonstrate the ability to think in terms of systems and apply systems thinking practices in archival work, including investigating a complex system as a whole, understanding the relationships within a system, thinking in levels, communicating information about a system, and defining systems and managing complexity.

Records/Data

2. Graduates are able to apply the foundational theories, methodologies and techniques of computational archival science to understand the effects of computing tools and methods on digital records and recordkeeping. Specifically, graduates have the ability to:

   2.1 demonstrate an ability to understand and apply foundational computational archival science theories to gain insights into the nature and evolving forms of digital records and recordkeeping.

   2.2 understand and critically evaluate the application of computational techniques and tools, such as those used in processes of "datafication", in relation to their impact upon the trustworthiness of records, such as their accuracy, reliability and integrity.

   2.3 understand and critically evaluate the impact of applying computational techniques and tools in relation to the selection and acquisition of records for preservation.

   2.4 understand and critically evaluate the impact of applying computational techniques and tools to sensitive information and its consequences for privacy.

   2.5 demonstrate an understanding and ability to conduct risk assessments of threats and vulnerabilities to archival cyberinfrastructures and records and develop risk mitigation plans and strategies.

Social

3. Graduates are able to appreciate foundational theories, methodologies and techniques of computational archival science to understand the effects of computing tools and methods on society. Specifically, graduates have the ability to:

   3.1 demonstrate an ability to critically examine the discursive aspects of data as well as the inherent power dynamics and cultural positionality manifest within computational infrastructure and the logics of computing.

   3.2 assess the social impact and usability of emerging technologies for archival purposes.

   3.3 draw upon knowledge of law and ethics in relation to all functions and activities related to applying computing technology to records and archives.

   3.4 engage critically and creatively with emerging professional and societal issues, such as bias in training data, arising from the application of computing technology to archival work

   3.5 advocate for the responsible and ethical application of computing techniques and tools in archival work.

We hope that programs offering archival education can use these proposed competencies to develop the programmatic CAS learning outcomes and curricular content that best speaks to their students and specific contexts. As more archival programs engage in teaching CAS, we believe that it would be beneficial for there to be a period of broader discussion facilitated by archival professional associations about the content of CAS graduate competencies and how best to align, or integrate, them with existing guidelines on graduate archival education.

A final point about pedagogy: our experience indicates that CAS requires students to "get their hands dirty" with the technology to truly understand it. This implies that students should learn by doing lab-based in-class activities and group or individual assignments that help to build their experience and skills in working with and applying technology to a range of archival tasks. We would add that this does not require archival educational programs to have access to, or create, advanced computing facilities. We have managed to teach CAS even using only free online tools and libraries.

5. Conclusion

Computing technology has revolutionized archival work, starting with the adoption of digital records and evolving with emerging technologies like AI and blockchain. This transformation includes digitization for easier access and preservation, applying advanced computational methods for data analysis, and tackling challenges in managing and preserving digital records. The rise of Computational Archival Science reflects these changes, focusing on computational tools and methods to improve archival practices and address emerging challenges in digital recordkeeping. Computational Archival Science (CAS) is still nascent and evolving, but it emphasizes integrating computing with archival science while advocating transdisciplinary collaboration. It also adopts hands-on learning, aiming to develop competencies across technical, record-keeping, and societal dimensions in archival education and the application of these competencies in archival work. As the case studies featured in this chapter illustrate, it is possible to teach CAS both to students with arts backgrounds and engineering backgrounds; in fact, to advance the development of CAS, pedagogy must be designed to support both groups. Additionally, CAS does not have to be at odds with sensitivity to recent archival scholarship emphasizing postcolonialism and progressive themes – as the many CAS papers focusing on ethical and social justice aspects of the intersection of computing and archival work demonstrate, invalidating what. we referred to in the abstract as a false dichotomy.

The teaching of CAS is not merely an option but a necessity for the archival profession to stay relevant and responsive to the changing landscape. It will equip professionals with the tools and perspectives needed to navigate the complexities of digital recordkeeping, ensuring that archives remain accessible, trustworthy, and reflective of our evolving society. This commitment to evolving education in CAS highlights its significance in shaping a future where archives and the archival perspective are more than applicable to the past, but dynamic resources and ways of understanding and engaging with our present world.

Acknowledgments

On the U.S. side, we wish to acknowledge the support of three IMLS (Institute of Museum and Library Services) grants: (1) "Computational Thinking to Library and Archival Science Education & Research, CT-LASER" (RE-73-18-0105-18) (2018-2020: link), (2) "Piloting an Online Collaborative Network for Integrating CT into Library and Archival Education and Practice, Piloting Network" (RE-246334-OLS-20) (2020-2024: link), and (3) "**T**raining of **A**rchival & **L**ibrary **E**ducators with i**N**novative **T**echnologies, TALENT Network" (RE252287-OLS-22) (2022:2025: link). This work benefited from the contributions of Dr. Michael J. Kurtz, RIP Dec. 17, 2022, who was an inspiring mentor, collaborator, and dear friend of many in this CAS research community.

On the Canadian side, we acknowledge the support of **Boeing Canada** in developing training for archival and LIS studies students in visualization and visual analytics and Canada's **Natural Science and Engineering Research Council**'s funding for a Collaborative Research and Education Training Experience (CREATE) grant (CREATE 528125-1) to establish the Blockchain Graduate Pathway.

References

Baron, J.R., Sayed, M.F. and Oard, D.W. (2022). Providing more efficient access to government records: A use case involving application of machine learning to improve FOIA Review for the deliberative process privilege. *Journal on Computing and Cultural Heritage* (JOCCH) 15(1), 1–19. https://dl.acm.org/doi/10.1145/3481045.

Buchanan, S. A., Gracy, K. F., Kitchens, J. & Marciano R. (2022). Computational Thinking Integration into Archival Educators' Networked Instruction. In 2022 IEEE International Conference on Big Data, 7th CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2022/12/Sarah_Buchanan.pdf.

Buchanan, S. A., Wachtel J. L., & Stevenson J. A. (2023). Accelerating Precision Research and Resolution Through Computational Archival Science Pedagogy. In 2023 IEEE International Conference on Big Data, 8th CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2023/11/S01204_4947.pdf.

Coindesk (2023, May 11). Top Blockchain University: University of British Columbia. *Coindesk. https://www.coindesk.com/learn/top-blockchain-university-university-of-british-columbia/*

Gilliland, A. (2023). Preparing Archivists in Computational Thinking and Innovative Technologies. International Academic Conference on Digital Intelligence: Empowering the Modernization of Archival Work. School of History and Culture, Shandong University, China. https://www.youtube.com/watch?v=ZdLJHQLbR4k.

Graham, S., Milligan, I., Weingart, S. B., & Martin, K. (2016). *Exploring big historical data: the historian's macroscope.*

Greenberg, J. et al. (2023). IMLS-funded LEADING Project. https://mrc.cci.drexel.edu/leading/.

Jaillant, L. (2022). *Archives, access and artificial intelligence: working with born-digital and digitized archival collections*. Bielefeld University Press.

Jaillant, L. & Aske, K. (2024). Are Users of Digital Archives Ready for the AI Era? Obstacles to the Application of Computational Research Methods and New Opportunities. *Journal on Computing and Cultural Heritage* (JOCCH) 16(4), Article 87. https://doi.org/10.1145/3631125

Jo, E. S., & Gebru, T. (2020). Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 306-316). ACM Digital Library.

Lemieux, V. L., & Feng, C. (2021). Conclusion: Theorizing from Multidisciplinary Perspectives on the Design of Blockchain and Distributed Ledger Systems (Part 2). In *Building Decentralized Trust: Multidisciplinary Perspectives on the Design of Blockchains and Distributed Ledgers* (pp. 129-163.). Springer International Publishing.

Lemieux, V., Hofman, D., Batista, D., & Joo, A. (2019). Blockchain technology & recordkeeping. ARMA *International Educational Foundation*. https://armaedfoundation.org/wp-content/uploads/2021/06/AIEF-Research-Paper-Blockchain-Technology-Recordkeeping.pdf

Lemieux, V., Voskobojnikov, A., & Kang, M. (2021, July). Addressing audit and accountability issues in self-sovereign identity blockchain systems using archival science principles. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference* (COMPSAC) (pp. 1210-1216). IEEE.

Ludäscher, B. (2016). A brief tour through provenance in scientific workflows and databases. In *Building trust in information: Perspectives on the frontiers of provenance* (pp. 103-126. Springer International Publishing.

Marciano, R., Lemieux, V., Hedges, M., Esteva, M., Underwood, W., Kurtz, M. & Conrad, M. (2018a). Archival records and training in the Age of Big Data. In J. Percell , L. C. Sarin , P. T. Jaeger , J. C. Bertot (Eds.), *Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education* (Advances in Librarianship), Volume 44B (pp.179-199). Emerald Publishing Limited. https://ai-collaboratory.net/wp-content/uploads/2020/10/Marciano-et-al-Archival-Records-and-Training-in-the-Age-of-Big-Data-final.pdf

Marciano, R., Underwood, W., Hanaee, M., Mullane, C., Singh, A., Tethong, Z. (2018b). Automating Detection of PII in WWII Camp Records. 2018 IEEE International Conference on Big Data, 3rd CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2020/03/2.Marciano.pdf

Marciano, R. Aggarat, S., Frisch, H., Hunt, M.R., Jain, K., Kocienda, G., Krauss, H., Liu, C., McKinley, M., Mir, D., Mullane, C., Patterson, E., Pradhan, D., Santos, J., Schams, B., Shiue, H.S.Y., Silva, A.J., Suri, M., Turabi, T., Vaselli, M., & Xu, J. (2019a). Reframing Digital Curation Practices through a Computational Thinking Framework. In 2019 IEEE International Conference on Big Data, 4th CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2020/04/ReframingDC-UsingCT_final.pdf.

Marciano, R., Lemieux, V, Hedges, M., Tomiura, Y., Katuu, S., Greenberg, J., Underwood, W., Fenlon, K., Kriesberg, A., Kendig, M., Jansen, G., Piety, P., Weintrop, D., & Kurtz, M. (2019b). Establishing an International Computational Network for Librarians and Archivists. R., V. Lemieux, V., M. Hedges, Tomiura, Y., Shadrack, K., Greenberg, J., Underwood, W., Fenlon, K., Kriesberg, A., Kendig, M. In iConference 2019 Blue Sky Papers series, IDEALS Institutional Repository. http://hdl.handle.net/2142/103139.

Marciano, R. (2021). AFTERWORD: Towards a new Discipline of Computational Archival Science (CAS). In Jaillant, L (Ed.). *Access and Artificial Intelligence: Working with Born-Digital and Digitised Archival Collections*, Bielefeld University Press. https://www.transcript-open.de/doi/10.14361/9783839455845-009?html#read-container.

Milligan, I. (2019). *History in the age of abundance?: How the web is transforming historical research*. McGill-Queen's University Press.

Mordell, D. (2019). Critical questions for archives as (big) data. *Archivaria*, 87(87), 140-161. https://archivaria.ca/index.php/archivaria/article/view/13673.

National Archives (2016). The Application of Technology-assisted Review to Born-digital Records Transfer, Inquiries and Beyond. http://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf

PACG (2011) Educating for the Archival Multiverse. *The American Archivist* 74, pp.69-101.

Payne, N. (2018). Stirring The Cauldron: Redefining Computational Archival Science (CAS) for The Big Data Domain. In 2018 IEEE International Conference on Big Data, 3rd CAS Workshop, https://ai-collaboratory.net/wp-content/uploads/2020/03/4.Payne_.pdf

Piety, P., Conrad, M., Marciano, R., Cornfield, I., Dallimore, E., Fettig, R., Hansen, E., Kemp, H., Turabi, T. (2023). Teaching and Learning with Archival Materials through the Development of Interactive Computational Notebooks. In Denison, V., Kearns, S. K., Leimkuehler, R., and Rogova, I. (Eds.), 2024 *Archives and Primary Source Handbook*. New Prairie Press. https://ai-collaboratory.net/wp-content/uploads/2023/10/Piety_Conrad_Marciano_et_al-FINAL.pdf.

Suleman, Z. (2022). *Learning to trust: exploring the relationship between user engagement and perceptions of trustworthiness in self-sovereign blockchain systems* (Master's dissertation, University of British Columbia). https://open.library.ubc.ca/soa/circle/collections/ubctheses/24/items/1.0417415?o=1

Suleman, Z. J., & Lemieux, V. L. (2023). Learning to Trust: Exploring the Relationship between Trust and User Experience in Blockchain Systems. In *Blockchain in Healthcare: Analysis, Design and Implementation* (pp. 119-144). Springer International Publishing.

Underwood, W., Weintrop, D., Michael Kurtz, M., Marciano, R. (2018). Introducing Computational Thinking into Archival Science Education. In IEEE International Conference on Big Data, 3rd CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2020/03/1.Underwood.pdf.

Underwood, W. (2019a). Computational Thinking in Archival Science Research. Presentation at the Developing a Computational Framework for Library and Archival Education Workshop, iSchool Conference. https://ai-collaboratory.net/wp-content/uploads/2020/04/Underwood_CompThinkInArchResearch.pdf, with final report at: https://ai-collaboratory.net/wp-content/uploads/2020/11/Final_Report_r.pdf

Underwood, W. Marciano, R. (2019b). Computational Thinking in Archival Science Research and Education. In 2019 IEEE International Conference on Big Data, 4th CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2021/03/Underwood.pdf.

Underwood, W. Gage J. (2023). Can GPT-4 Think Computationally about Digital Archival Practices? In 2023 IEEE International Conference on Big Data, 8th CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2023/11/S01213_7999.pdf.

Underwood, W. Gage J. (2024). Training in Computational Archival Science: Do CAS Educational Frameworks meet Professional Expectations? In 2024 IEEE International Conference on Big Data, 9th CAS Workshop. https://ai-collaboratory.net/wp-content/uploads/2024/11/S01206_4692.pdf.